Distributionally Robust and Structure Exploiting Algorithms for Power System Optimization Problems

A dissertation submitted to

The University of Liverpool

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

By

Chao Duan

Supervisor: Dr. Lin Jiang

Electrical Engineering & Electronics

October 2018

Acknowledgements

I cannot overstate my gratitude to my supervisor Dr. Lin Jiang who makes my liverpool life so enjoyable and rewarding. His invaluable guidance, encouragement, and advice have helped me go through the most difficult times during my Ph.D. study. The unsurpassed academic environment he provided broadens my research ideas and makes my Liverpool experience indispensable and unforgettable.

Also, my deep gratitude goes to Prof. Wanliang Fang, who is not only my supervisor but also life mentor. In the past six years, his insightful teaching, patient guidance, and helpful advice have helped me tremendously throughout my journey toward Ph.D. The high academic taste and rigorous scholarship of him have shaped my views on research and will continue to influence my academic career. He not only academically inspires but also philosophically enlightens me, which impacts my views on the society, value, life, and happiness.

I would like to thank Dr. Jun Liu for his constant accessibility, friendship and helpful suggestions for various problems I encountered during my study. The discussions with him have always been enjoyable and beneficial to me. My sincere thanks also go to Dr. Chuanke Zhang. The interaction with him has helped me enter a new research field of stability and control of time-delay systems. I would like to extend my gratitude to Dr. Chongtao Li. The stimulating discussions with him have broadened my research ideas and he kindly provided the codes of the benchmarking approach in Chapter 9 of this thesis. My deep thanks also go to Prof. Zhaohong Bie, Prof. Jianxue Wang, Prof. Dengfu Zhao and Prof. Zhengchun Du for their courses of power system analysis that open up to me the wonderful field of power systems to which I decide to devote the rest of my life.

I am grateful to my friends and colleagues at Xi'an and Liverpool: Zhixiang Wang, Zhanhong Wei, Danhua Mei, Yongqian Yang, Xi Nie, Xiaojuan Zhou, Yang Liu, Kai Shi, Qi Zhu, YueFang Du, Yingjie Wang, Li Yao and many others for making life more excited and colorful.

I am deeply in debt to my parents and family for their unconditional love, support, and encouragement throughout all these years. My father and mother's belief in me has supported me through difficult times of my life. My deepest gratitude goes to my wife for her love, support, understanding, and tolerance for my staying away from home. I warmly dedicate this thesis to her.

I would like to thank the International PGR Programme of the University of Liverpool, the National Key Research and Development Program of China (2016YFB0901903), and China Scholarship Council for financial support.

Title: Distributionally Robust and Structure Exploiting Algorithms for Power System Optimization Problems
Applicant: Chao Duan

ABSTRACT

The modern power systems are undergoing profound changes as the large-scale integration of renewable energy and increasingly close interconnection of regional power grids. The intermittent renewable sources are bringing significant uncertainties to system operation so that all the analysis and optimization tools for the power system steady-state operation must be able to consider and manage the uncertainties. The large-scale interconnection of power systems increases the difficulty in maintaining the synchronization of all generators and further raises the challenging problem of systematically design multiple local and wide-area controllers. In both steady-state and dynamical problems, the large-scale interconnection is increasing the problem scale and challenging the scalability of analysis, optimization and design algorithms. This thesis addresses the problems of power system operation optimization under uncertainties and control parameter optimization considering time delays. The contributions are as follows.

This thesis proposes data-driven distributionally robust models and algorithms for unit commitment, energy-reserve-storage co-dispatch and optimal power flow problems based on novel ambiguity sets. The problem formulations minimize the expected operation costs corresponding to the worst-case distribution in the proposed ambiguity set while explicitly considers spinning reserve, wind curtailment, and load shedding. Distributionally robust chance constraints are employed to guarantee reserve adequacy and system steady-state security. The construction of ambiguity set is data-driven avoiding presumptions on the probability distributions of the uncertainties. The specific structures of the problem formulation are fully exploited to develop a scalable and efficient solution method.

To improve the efficiency of the algorithms to solve the operation and control optimization problems, this thesis investigates computational techniques to exploit special problem structures, including sparsity, chordal sparsity, group symmetry and parallelizability. By doing so, this thesis proposes a sparsity-constrained OPF framework to solve the FACTS devices allocation problems, introduces a sparsity-exploiting moment-SOS approach to interval power flow (IPF) and multi-period optimal power flow (MOPF) problems, and develops a structure-exploiting delay-dependent stability analysis (DDSA) method for load frequency control (LFC).

The power system stabilizers (PSS) and FACTS controllers can be employed improve system damping. However, when time delays are considered, it becomes more difficult to analyzing the stability and designing the controllers. This thesis further develops time-domain methods for analysis and synthesis of damping control systems involving time delays. We propose a model reduction procedure together with a condition to ensure the ϵ-exponential stability of the full-order system only using the reduced close-loop system model, which provides a theoretical guarantee for using model reduction approaches. Then we formulate the damping control design as a nonlinear SDP minimizing a carefully defined H_2 performance metric. A path-following method is proposed to coordinately design multiple damping controllers.

CONTENTS

1 Introduction

1.1 Background

Electric power systems are the most clean and effective way to transmit and distribute energy. All other forms of energy, including fossil fuel, hydro, wind, solar, and nuclear etc., can be conveniently converted into electric energy and fulfill various needs of the end-users. At the consumer side, the utilization of electric energy is much more efficient than directly using other forms of energy, e.g. the energy efficiency of a diesel vehicle is only 60% of that of an electrical vehicle. More importantly, the end-user usage of electricity is completely emission free. Therefore, the electric power systems have been and will continue to be the backbone infrastructure for human society. Currently, the modern power system is undergoing far-reaching changes from different aspects among which we will focus on the large-scale integration of renewable energy and large-scale interconnection of power systems.

In the past two and a half century, the fossil fuels, mainly the coal, oil and gas, have supported the development of the industrial civilization. After the large-scale exploitation and utilization since the Industrial Revolution, there will inevitably be an exhaustion of fossil fuels due to their non-renewable nature. The overexploitation of fossil fuels has also brought environmental problems, including air pollution and climate change, which threatens the sustainability of human society. In stark contrast to the unsustainability and environmentally unfriendly of fossil fuels, the hydro, wind and solar energy are clean, renewable and extremely abundant around the global. According to the estimation of World Energy Council, the total amount of renewable energy that could be exploited and utilized is 1.5×10^8 GW·h per year which amounts to 38 times of all the explored fossil energy. Such a large amount of energy is far beyond the current usage of human society. Therefore, shifting from fossil fuels to renewable energy is an irresistible trend and constitutes the energy revolution we are currently undergoing. From 2006 to 2016, the wind power global capacity increased from 74 GW to 487 GW, and the solar PV global capacity increased from 6 to 303 GW [1]. In 2016, the world installed more renewable power capacity than the capacity added from fossil fuels. By the end of 2016, renewables comprised an estimated 30% of the world's power generating capacity. All the renewables are integrated into the electric power systems, transmitted through the power networks and then distributed to consumers.

Large-scale integration of renewable energy brings challenges to the operation of power systems due to the uncertain and intermittent nature of renewable generation. In the conventional power systems, the generation side is completely controllable and the uncertainties only arise from the demand side. But when the renewables penetrate the system at a non-negligible level, both the generation and demand sides of the power system are subject to uncertainties.

Compared to the load demand, the renewable generation is more variable and unpredictable. For instance, the standard deviation of hourly aggregated wind generation is 17.9% of the installed capacity in Nordic and Baltic countries in 2014, and the probability of an hourly generation below 8% of the capacity is around 0.05 [2]. Considering the fact that the geographical dispersion of the renewable generation in Nordic and Baltic countries has reduced the uncertainties, the situation could be more severe in countries like China where large-scale centralized wind and PV power plants are constructed. As a result, all the computational tools for power system operation, including power flow analysis, unit commitment, economic dispatch and optimal power flow etc. need to take the variability and uncertainties into consideration and also incorporate new technologies and strategies like energy storages and demand-side response.

The electric energy still cannot be stored in a very large scale. The generation, transmission and distribution must be completed simultaneously. The electric energy supply and demand are always in real-time balance. In addition, the energy supply centers are geographically far away from the demand centers in many parts of the world. For example, in China, the eastern coastal regions produces 61.7% of the country's GDP, consumes 57.1% of the total electricity, but only 10.5% of the coal and 7.3% of the hydro resources locate in these areas [3]. In addition, 6 of the 8 largest wind power bases with each capacity larger than 10 GW locate in the less developed western and northern parts of the country. A similar situation appears in other countries. Brazil has very rich hydro resources especially in the Amazon river system located in the less populated North, while the load and economic centers are in the south and southeast of the country. Moreover, the Arctic circle contains 20% of the total wind power resources of the world but it is very sparsely populated. In all the cases listed above, the large-scale interconnection of power systems is almost the only choice toward better utilization of the natural energy resources. From the viewpoint of power engineering, the large-scale interconnection of power systems has the following advantages. First, the interconnection of power grids provides the physical basis for the optimal dispatch of the electricity generated from different primary energy resources dispersed in a large geographical area, which improves the economic efficiency of the gross energy utilization. Second, The aggregation of load demand and renewable generation in different areas could alleviate the variability and uncertainties of the net demand thereby reducing the spinning reserve required for safe operation of the system. Third, large-scale interconnection can also reduce the required contingency reserve, improve system reliability and increase system inertia. Due to the above reasons, interconnected power systems have been constructed in China, Europe, North American, Brazil and some other parts of the world. In recent years, the State Grid Company of China is even promoting the idea of global energy internet and the Global Energy Interconnection [4] Development and Cooperation Organization was established in Beijing in 2016.

The large-scale interconnection of power systems also brings various challenges to power

system analysis, optimization and control. The most immediate result of interconnection is the increase of problem scales. Numerical algorithms for power system operation and planning therefore must have enough ability to scale up to real-world large-scale problems. This aspect is especially true when large-scale interconnection interacts with large-scale introduction of uncertainties. Many algorithms for stochastic optimization of power systems suffer from the statistical curse of dimensionality and/or computational curse of dimensionality. On the other hand, the interconnected power systems also face typical stability problems. The AC power system is an electromechanical system whose safe operation relies on the synchronization of connected generators. The synchronized state is always disturbed by the continuous load/generation changes and sporadic faults. When the system becomes more and more interconnected, the oscillation mode between generators in one area against those in another area could be easily excited, which results in low-frequency oscillation across the system if the damping control systems are not carefully designed and implemented. To improve the observability and situation awareness of large-scale interconnected power systems, wide area measurement systems (WAMS) have been quicked developed and applied to real-world applications. WAMS provides the infrastructure for realizing the idea of wide-area damping control which can suppress the inter-area oscillations much more effectively than the conventional power system stabilizers using local feedback signals.

The trends of large-scale integration of renewable energy and large-scale interconnection of power systems are the most relevant undergoing changes to this thesis. Other trends of the modern power system may include the increasingly high penetration of power electronics, the ever-growing integration of electric power systems with other energy systems, and the restructuring of the business and operation models, etc. All these profound changes will have significant technological, economic, environmental as well as social impacts on human society. These undergoing changes also open up for new research opportunities.

1.2 Research Status-quo and Challenges

To facilitate the large-scale integration of renewable energy and large-scale interconnection of power grids, the following questions need systematical answers. 1) How do we deal with the uncertainties in power flow analysis and operational optimization? 2) How can we develop efficient algorithms to solve large-scale problems in power system optimization, stability analysis and control design? 3) How can we consider the imperfection of communication networks in stability analysis and design of damping control systems involving wide-area measurements? 4) How can we design the local and wide-area controllers in an automatic and coordinated manner to improve system damping? In the rest of this section, we provide a review of the-state-of-art in answering some of the questions listed above.

1.2.1 Power Flow Analysis under Uncertainties

Power flow analysis methods in the presence of uncertainty can be classified into two groups according to how uncertainty is represented. The first group is the probabilistic power flow (PPF) in which loads and generations are expressed as random variables with associated distribution functions. PPF aims at deriving the probability distribution of power flow solutions. Monte Carlo simulation based techniques [5, 6], analytical methods [7–9] and point estimate methods [10, 11] are extensively investigated to deal with this problem. By assuming the uncertain nodal power injections to obey pre-defined probability distribution, PPF aims at obtaining the probability distribution of power flow responses, including bus voltages and line flows. Vast literature has been devoted to developing solution method to PPF problem [5–14]. Monte Carlo simulation (MCS) [5, 6] obtains PPF solution by repeatedly running deterministic power flow. It is usually regarded as the benchmark method yet criticized for its huge computational burden. To improve computational efficiency, analytical methods including conventional convolution [7, 12, 13] and cumulants method [8, 9, 14] have been proposed to obtain approximate solution of PPF. These methods usually linearize the AC power flow equations at the point of expected power injections to express the power flow response as the linear combination of power injections. Based on the linear formulation, conventional convolution directly obtain the probability distribution through convolution process, whereas cumulants method first obtains the cumulants of the power flow responses then estimate probability distribution using series expansions. Point estimation [10, 11] is another approximate method. It estimates the moments of the power flow responses by running deterministic power flow at selected points. The probability distribution can also be constructed using series expansions based obtained moment information. Recently, stochastic response surface method [15] has also been applied to PPF problem. It models the power flow responses as polynomial chaos expansion with standard random variables and determines the unknown coefficients by running several deterministic power flows.

The second group of methods, interval power flow (IPF) [16–20], model uncertainty of loads and generations as intervals without distribution structures, which seems a more practical approach because system operators need to consider the worst-case scenario to guarantee the security of the power networks. Interval arithmetic (IA) is introduced in [16] to modify conventional Newton iterations to obtain the outer approximation of the solution set, but the results are highly conservative and the convergence is unfounded. A local search procedure is proposed in [17] to find accurate boundary solutions but the convergence is not proved and only local optimality is guaranteed. To overcome the limitation of the IA method, affine arithmetic (AA) is used in [18] by expressing power system state variables as an affine function of uncertainties, and linear programming (LP) is employed to obtain tight bounds of uncertainty

parameters. However, this method is still approximate in nature. In [19], from a range arithmetic perspective, a nonlinear programming (NLP) model with complementarity constraints is proposed to find the upper and lower bounds of power flow solutions. Although the formulation is non-convex and difficult to solve, this work provides a theoretical foundation for optimization-based method to solve IPF. In the most recent work [20], IPF is formulated as a quadratically constrained quadratic programming (QCQP) problem. Convex envelops are employed to relax the non-convex QCQP to a convex LP. Starting from estimated solution intervals, the optimality-based bounds tightening (OBBT) method is introduced to obtain the tight outer approximation of the feasible region. Though this method can obtain less conservative interval solutions than previous methods, the results in any case are still supersets of the exact intervals.

1.2.2 Power System Optimization under Uncertainties

We are not intended to review all aspects of power system optimization under uncertainties. Our purpose here is to exposit the key ideas and methods to deal with uncertainties in power system optimization problems. Focuses are given to four approaches, namely scenario-based stochastic program (SBSP), chance-constrained program (CCP), robust optimization (RO), and distributionally robust optimization (DRO).

SBSP approach assumes operational uncertainties, i.e. load forecasting inaccuracies and uncertain renewable generation, follow pre-defined probability distributions which can be obtained from historical data. The objective is to minimize the expected value of generation costs. Two-stage problem formulations share the form

$$\min_{\boldsymbol{y}\in\Xi}\left\{G(\boldsymbol{y})+\min_{\boldsymbol{x}\in\Omega^{\mathbb{P}}(\boldsymbol{y})}\mathbb{E}_{\mathbb{P}}\{F(\boldsymbol{x},\boldsymbol{\xi})\}\right\} \tag{1-1}$$

where $\boldsymbol{y}$ denotes the first-stage commitment decision variables of generation units, $\boldsymbol{x}$ denotes the second-stage dispatch variables and $\boldsymbol{\xi}$ is a random vector representing uncertainties. Since continuous distributions are usually numerically intractable, they are replaced by discrete scenarios. So the objective becomes minimizing the weighted-average generation cost over parallel scenarios for two-stage problems [21, 22] or a scenario tree for multi-stage problems [23–25]. Thus the solution quality of SBSP relies on the representativeness of selected scenarios. Monte Carlo simulation based method is usually employed to generate scenarios [22, 25]. A huge number of scenarios are often required to comprehensively represent the underlying stochastic nature [26], which results in prohibitively high computational burden [26, 27]. Therefore, SBSP is often equipped with scenario reduction methods to control computational complexity [25, 26]. However, due to the approximate nature of SBSP, the dilemma between quality and complexity of scenarios always exists. Consequently, the robustness of the scheduling strategy obtained cannot be rigorously guaranteed.

CCP approach is motivated by the fact that SBSP usually requires all technical operational constraints being satisfied under any scenarios, which may include extremely rare events and increase cost unnecessarily [27]. In CCP, under pre-defined probability distribution of the uncertainties, some constraints are required to be met with a specified high probability, which is written in the general form

$$\mathbb{P}\{L(\boldsymbol{\xi}) \leq 0\} \geq 1 - \varepsilon \tag{1-2}$$

where $L(\cdot)$ is a function determined by other decision variables representing certain technical constraint. In [28], chance constraints are applied to load satisfaction. Under normal distribution, such chance constraints are transformed to deterministic linear equalities with some conservatism. In [22], chance constrained are employed to ensure the utilization of wind power, and such problem is solved by a Monte Carlo simulation based sample average approximation (SAA) method. Since the transformation in [28] is restrictive and conservative and SAA used in [22] is approximate in nature, there exist no elegant methods to equivalently convert the chance constraints to deterministic constraints to facilitate optimization computation.

In stark contrast to SBSP and CCP, RO does not require any probability distribution information of the uncertainties. Instead, randomness is represented by a deterministic uncertainty set which includes worst-case scenarios. The objective is minimizing the worst-case cost regarding all possible realizations of the uncertainty set. The following two-stage formulation attracts most attention [29–33]:

$$\min_{\boldsymbol{y}\in\Xi} \left\{ G(\boldsymbol{y}) + \max_{\boldsymbol{\xi}\in\mathcal{D}} \min_{\boldsymbol{x}\in\Omega(\boldsymbol{y})} F(\boldsymbol{x}, \boldsymbol{\xi}) \right\} \tag{1-3}$$

where $\mathcal{D}$ denotes the uncertainty set. The numerical tractability and solution quality of above formulation both rely on the design of uncertainty set $\mathcal{D}$ which are usually defined by a set of linear inequalities. To solve the second stage max-min. problem in (1-3), the innermost min. problem needs to be dualized so that the max-min. problem is transformed into a non-convex bilinear max. problem which is NP-hard in general. This bilinear problem can be either solved by outer approximation method [30] (only local optimality is guaranteed) or rewritten into a mixed integer linear program (MILP) using big-M method [29, 32] (under some assumption on $\mathcal{D}$, see [32]) and solved by generic MILP solvers. Although RO provides security against worst-case scenario, it may also yield over-conservative solutions resulting from the sheer ignorance of underlying probabilistic information.

Some works attempt to melt the bright sides of above approaches and make remedies to their disadvantages. In [22], chance constraints are integrated into the two-stage SBSP framework to improve wind power utilization. Reference [34] proposes a UC (unit commitment) formulation that combines SBSP and RO to achieve lower expected cost while ensuring system robustness. In [33], risk consideration on specific scenarios is incorporated into two-stage

RO.

Note that the SP assumes the underlying probability distribution of uncertainties to be precisely known, whereas RO ignores the probabilistic information. In practice, the probability distribution truly exists but must be estimated from historical data and is, therefore, itself uncertain. To better modeling and tackling uncertainties, in several very recent studies, distributionally robust optimization (DRO) has been introduced to power system optimization problems including unit commitment [35, 36], energy reverse dispatch [37, 38], reverse scheduling [39, 40] and DC optimal power flow [41, 42]. DRO assumes that the true probability distribution of uncertain parameters lies in an ambiguity set (of probability distributions) and immunizes the operation strategies against all distributions in the ambiguity set. Different ways to construct the ambiguity set leads to different DRO approaches with the different degree of conservatism and computational efficiency. Paper [35] adopted a scenario-based approach where the random variable representing wind generation is assumed to have finite support. Statistical inference technique is employed to construct ambiguity sets for the discrete distributions. This method is data-driven and more data leads to the less conservative solution. Other papers assume that the random variables have continuous distributions. In [36], the support of a one-dimensional random variable is partitioned into several segments, and the ambiguity set imposes an upper bound for the expectation in each segment. As the number of segments increases, the probability distribution can be characterized by more details. The ambiguity sets employed in [37–40, 42] are the sets of all probability distributions with given mean and covariance. Paper [38] further assumes unimodality to reduce conservatism, and reference [42] further considers the uncertainties of mean and covariance. Different from all others, paper [41, 43] assumes the distribution type is known as a priori while the mean and covariance are subject to uncertainties.

1.2.3 Time-delay Power System Stability Analysis and Control Design

Two groups of methods have been applied to analyze the stability and performance of time-delay load frequency control (LFC) and wide-area damping control （WADC）. The first group is the frequency-domain method based on eigenanalysis [44–47]. To obtain the eigenvalues, the time-delay system is often approximated by an augmented delay-free linear system through discretization of the continuous state function of the time-delay system. Usually, the dimension of the delay-free system is much higher than the time-delay system and proportional to the number of discretization points employed. As the number of discretization points N increases, the eigenvalues of the delay-free system converge to those of the time-delay system at the rate of $\mathcal{O}(N^{-N})$. Hence, practically speaking, frequency-domain methods can obtain the exact eigenvalues along with other accurate performance metrics, e.g. delays margins, damping ratios and damping factors, for systems with constant delays. The classi-

cal frequency-domain phase compensation approaches [48–50] have been the most practically applicable methods to design PSSs and FACTS controllers to suppress oscillations. However, their extensions to time-delay power systems are scarcely reported.

The second group of methods is the time-domain delay-dependent stability analysis [51–55] which leverage linear matrix inequalities (LMI) as the computational tools. In time-domain methods, the stability of the time-delay systems is certified by the existence of certain Lyapunov -Krasovskii functionals (LKF). To find such LKFs, a candidate LKF with undetermined weighting matrices is first constructed, and some integral inequalities are employed to bound the time derivative of the LKF. Then the positive definiteness of the LKF and the negative definiteness of its derivative are then expressed as LMIs. The time-domain methods are always conservative, but the conservatism can be reduced by using more general LKF and tighter integral inequalities. In addition, they can deal with both constant and time-varying delays and facilitate the robust design of the controllers. For example, LMI based delay-dependent stability criteria are employed to calculate the delay margins of WADCs installed at generator excitation systems [51] and FACTS devices [56]. Similar methods are also applied to LFC with constant and time-varying delays [57]. A more accurate and computationally cheaper LMI-based stability criterion is applied to multi-area LFC in [58]. Although time-domain methods possess some conservatism when calculating the delay margins, they can be conveniently applied to systems with single or multiple, constant or time-varying delays. Moreover, the development of the modern interior point method (IPM) provides off-the-shelf tools to solve LMIs.

The ultimate goal of stability analysis is to inform the controller design to mitigate the adverse effects of time delays. From this aspect, time-domain LMI based methods have advantages over frequency domain methods due to the fact that LMI based criteria can be conveniently employed to derive the delay-independent and delay-dependent bounded real lemmas (BRL) [59] which then can be used to design H_∞ robust controllers. For instance, delay-independent BRLs are used in [60] and [61] to design state-feedback and PI robust controllers for LFC, respectively. In [62], PID-type robust controllers are designed for LFC systems based on a delay-dependent BRL. In addition, delay-dependent BRLs are also employed in [53] to design state-feedback WADC. Above-mentioned controllers for LFC and WADC are all designed to minimize the H_∞ index while guaranteeing stability for any delays less than the preset upper bounds (the upper bounds are ∞ for delay-independent methods).

1.2.4 Remaining Challenges

Based on the overview of previous works, we have identified the following remaining challenges:

1. Although there have been several attempts to apply the idea of DRO to power system operation under uncertainties, the performance and efficiency of existing methods are

still not satisfactory. For instance, the DRO approach discussed in [38, 63, 64] has the following drawbacks. First, the ambiguity set characterized only by the first two moments is in fact very large thus the method is still very conservative. Second, moments also need to be estimated from historical data therefore are uncertain as well. Third, the problems are finally cast into a sequence of semi-definite programs which are very computationally intensive. The core issue here is the structure of the ambiguity sets employed in the DRO. How to construct an ambiguity set which is theoretically tractable, statistically reliable and numerically scalable is still a challenging problem.

2. The OPF and IPF problems are inherently non-convex optimization problems due to the non-linearity of power flow equations. The conventional interior point method can only produce a locally optimal solution. How to solve those non-convex problems to global optimality with acceptable computational resources is still challenging. Recent attempts [65–67] to convexify the OPF problem bring new insights into this field. The questions of how to construct ever-tighter convex relaxations with mild increase of computational burden and how to deal with the infeasibility and scalability issues of convex relaxations await systematical answers before the convex relaxation techniques can be applied to real-world problems.
3. The major obstacle of the LMI-based time-domain methods for time-delay power system stability analysis is the computational burden of solving large-scale LMIs due to the very limited ability of the state-of-the-art solvers to tackle large-scale problems. To resolve this problem, we have two pathways. The first one is to reformulate the problems into easier ones by exploiting the structural features of the power system control loops. The second one is to employ the model reduction methods, but the stability analysis result on the reduced-order model does not necessarily imply the same result for the original full-order system. How to ensure the stability of the full-order system by just using the reduced-order model is also a challenging problem.
4. The problem of coordinate design of multiple local and wide-area damping controllers also expect systematical solutions. Although it is quite easy for time-domain LMI-based methods to formulate the design problems due to its affinity to modern H_∞ and H_2 control theory. However, it is not always easy to actually solve those problems. Only when the full state feedback control is adopted, the problem can be equivalently formulated as convex semi-definite programs (SDP) thus solvable by off-the-shelf solvers [53]. For other control architectures, the problem is bilinear [68] or more general nonlinear SDPs which are non-convex and NP-hard.

1.3 Contribution and Thesis Outline

The research contents of this thesis can be summarized as in Fig. 1-1. The physical object studied in this paper is the power systems with the uncertain, interconnected and time-delay features. The basic theoretical tool is optimization theory and method with emphasis on the distributionally robust optimization. Meanwhile, focuses are given to exploiting special problem structures, e.g. sparsity and symmetry, to improve efficiency in the problem-solving processes. Specific power system applications considered include operation schedule, interval power flow, device allocation and control analysis & synthesis.

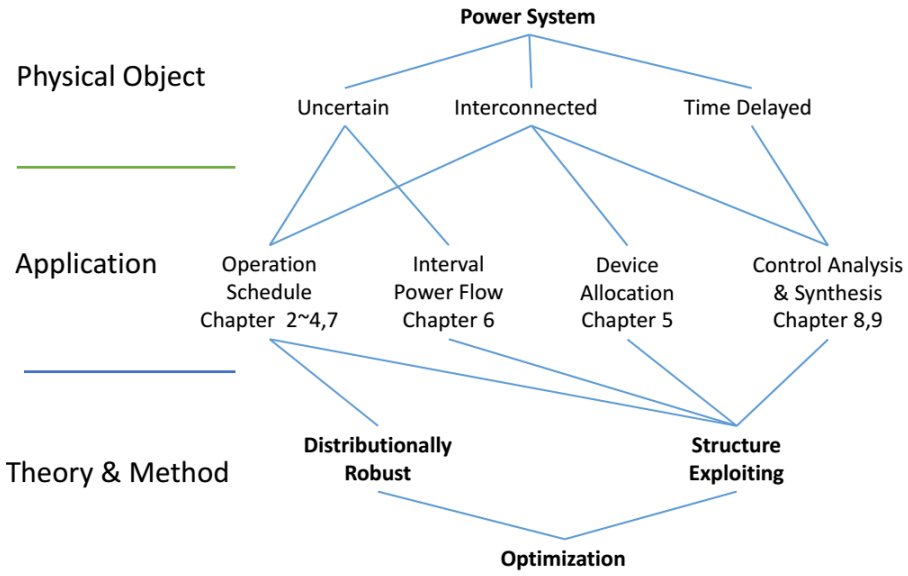


Figure 1-1 Summary of the Contents of this thesis

The contribution and organization of the rest of the thesis are as follows.

Chapter 2 proposes a data-driven affinely adjustable distributionally robust method for unit commitment considering uncertain load and renewable generation forecasting errors. The proposed formulation minimizes expected total operation costs, including the costs of generation, reserve, wind curtailment and load shedding, while guarantees the system security. Without any presumption about the probability distribution of the uncertainties, the proposed method constructs an ambiguity set of distributions using historical data and immunizes the operation strategies against the worst-case distribution in the ambiguity set. The more historical data is available, the smaller the ambiguity set is and the less conservative the solution is. The formulation is finally cast into a mixed integer linear programming whose scale remains unchanged as the number of historical data increases. Numerical results and Monte Carlo simulations on the 118- and 1888-bus systems demonstrate the favorable features of the proposed method.

Chapter 3 develops distributionally robust energy-reserve-storage co-dispatch model and

method to facilitate the integration of variable and uncertain renewable energy. The uncertainties of renewable generation forecasting errors are characterized through an ambiguity set which is a set of probability distributions consistent with observed historical data. The proposed model minimizes the expected operation costs corresponding to the worst-case distribution in the ambiguity set. Distributionally robust chance constraints are employed to guarantee reserve and transmission adequacy. More data also leads to smaller the ambiguity set and less conservative decision strategy. The formulation is finally cast into a mixed integer linear programming whose scale remains unchanged as the number of historical data increases. Inactive constraint identification and convex relaxation techniques are introduced to reduce the computational burden. Numerical results and Monte Carlo simulations on IEEE 118-bus systems demonstrate the effectiveness and efficiency of the proposed method.

Chapter 4 discusses a distributionally robust chance constrained approximate AC-OPF. The power flow model employed in the proposed OPF formulation combines an exact AC power flow model at the nominal operation point and an approximate linear power flow model to reflect the system response under uncertainties. The ambiguity set employed in the distributionally robust formulation is the Wasserstein ball centered at the empirical distribution. The proposed OPF model minimizes the expectation of the quadratic cost function w.r.t. the worst-case probability distribution and guarantees the chance constraints satisfied for any distribution in the ambiguity set. The whole method is data-driven in the sense that the ambiguity set is constructed from historical data without any presumption on the type of the probability distribution, and more data leads to smaller ambiguity set and less conservative strategy. Moreover, special problem structures of the proposed problem formulation are exploited to develop an efficient and scalable solution approach. Case studies are carried out on IEEE 14 and 118 bus systems to show the accuracy and necessity of the approximate AC model and the attractive features of the distributionally robust optimization approach compared with other methods to deal with uncertainties.

Chapter 5 formulates FACTS device allocation problem as a general sparsity-constrained OPF problem and employs $L_q(0 < q \leq 1)$ norms to enforce sparsity on FACTS devices setting values to achieve solutions with desirable device numbers and sites. An algorithm based on alternating direction method of multipliers is proposed to solve the sparsity-constrained OPF problem. The algorithm exploits the separability structure and decomposes the original problem into an NLP subproblem, an L_q regularization subproblem, and a simple dual variable update step. The NLP subproblem is solved by the interior point method. The L_q regularization subproblem has a closed-form solution expressed by shrinkage-thresholding operators. The convergence of the proposed method is theoretically analyzed and discussed. The proposed method is successfully tested on the allocation of SVC, TCSC and TCPS on IEEE 30-, 118- and 300-bus systems. Case studies are presented and discussed for both single-type and multiple-

type FACTS devices allocation problems, which demonstrates the effectiveness and efficiency of the proposed formulation and algorithm.

Chapter 6 describes a novel optimization-based method to obtain high-accuracy or even exact global solutions to IPF problems. At first, the IPF problems are formulated as polynomial optimization problems probably with rational objective functions. Then Lasserre's hierarchy, or moment-SOS (sum of squares) approach, is introduced to relax the non-convex problems to convex semidefinite programming (SDP) problems. Correlative sparsity in the polynomial optimization problems is exploited to improve numerical tractability and efficiency. Finally, case studies on IEEE 6-bus, 9-bus and 14-bus systems demonstrate the second-order moment relaxation is capable of obtaining exact global interval solutions on small-scale systems, and numerical results on IEEE 57-bus, 118-bus and 300-bus systems show the proposed method can significantly improve the interval solutions compared with recent Linear Programming (LP) relaxation method on larger systems.

Chapter 7 discusses a convex relaxation based decomposition algorithm to solve the full AC multi-period OPF with energy storages and renewable sources. Based on alternating direction method of multipliers (ADMM), the original time-correlated non-convex optimization problem is decomposed into two subproblems. The first subproblem is non-convex and separable among different time slots. Moment relaxation can be constructed and solved for each time slot in parallel. The second subproblem is a convex quadratic program (QP) and separable among different buses, which can be solved in parallel by standard interior point (IPM) solver. Case studies on a benchmark systems demonstrate the effectiveness of the proposed algorithm.

Chapter 8 investigates the computational aspect of delay-dependent stability analysis of LFC. The basic idea is to improve the numerical tractability of DDSA (delay-dependent stability analysis) by exploiting the chordal sparsity and symmetry of the graph related to LFC loops. The graph-theoretic analysis yields the structure restrictions of weighting matrices needed for the LMIs to inherit the chordal sparsity of the control loops. By enforcing those structure restrictions on weighting matrices, the positive semi-definite constraints in the LMIs can be decomposed into smaller ones, and the number of decision variables can be greatly reduced. Symmetry in LFC control loops is also exploited to reduce the number of decision variables. Numerical studies show the proposed structure-exploiting techniques significantly improves the numerical tractability of DDSA at the cost of the introduction of acceptable minor conservatism.

Chapter 9 addresses two important issues of the time-domain LMI-based approaches to analyze and design damping control for time-delay power systems. The first issue concerns with the deficiency of the reduced-order system model often used in the time-domain method to reliably reflect the stability of the original full-order system. Leveraging tools from dissipativity theory, we propose a model reduction procedure together with a condition to ensure

the ϵ-exponential stability of the full-order system only using the reduced close-loop system model, which provides a theoretical guarantee for using model reduction approaches. The second issue is the numerical intractability of the nonlinear SDP problems when attempting to coordinately design controllers. Instead of seeking global optimum of the problem, we propose a path-following method to systematically search for a local solution. The algorithm iteratively improves the damping factor and H_2 performance of the system by tuning the control parameters in an automatic and coordinated manner. Case studies on three benchmark systems demonstrate the design power the proposed method.

Appendix A details the interior point method we use in Chapter 5 of the thesis to solve the NLP sub-problem. Three measures, namely the adaptive barrier update strategy, the filter line-search method and the feasibility restore phase, are simultaneously introduced in the conventional primal-dual interior point method (IPM) framework to enhance the robustness of existing Optimal Power Flow (OPF) algorithms when applied to systems with considerable number of FACTS devices. Firstly, an adaptive barrier parameter strategy is employed to update the barrier parameter after the current μ-barrier problem solved to a certain accuracy. Secondly, a filter line-search procedure is introduced to generate the next iterate. Third, the algorithm initiates a feasibility restore phase as a remedy in case of getting stuck at a non-optimal point. Comparative case studies with previous algorithms on both standard test systems and large-scale real-world systems demonstrate the novel algorithm outperforms conventional IPMs in robustness and efficiency.

2 Data-driven Affinely Adjustable Distributionally Robust Unit Commitment

2.1 Introduction

Unit Commitment (UC) aims at reducing costs and improving reliability by optimal scheduling and dispatching generation units. Integration of intermittent renewable energy and market-driven operation have brought uncertainties to both generation and demand sides. It is, therefore, necessary to incorporate uncertainties into UC. Vast literatures are devoted to the stochastic optimization for UC [27]. Among all the methods proposed so far, the stochastic programming (SP) and the robust optimization (RO) attract the most attention.

SP approach assumes operational uncertainties follow pre-defined probability distributions which can be learned from historical data. The objective is to minimize the expectation of generation costs. Since continuous distributions are usually numerically intractable, they are replaced by discrete scenarios. Hence the objective becomes minimizing the weighted-average generation costs over parallel scenarios for two-stage problems [21, 22, 69] or a scenario tree for multi-stage problems [23–25]. The solution quality of SP relies on the representativeness of selected scenarios. Monte Carlo simulation based method is usually employed to generate scenarios [22, 25]. A huge number of scenarios are often required to comprehensively represent the underlying stochastic nature [26], which results in prohibitively high computational burden [26, 27]. Therefore, SP is often equipped with scenario reduction methods to control computational complexity [25, 26]. However, the dilemma between quality and complexity of scenarios always exists.

In contrast to SP, RO does not require any probabilistic information of the uncertainties. Instead, randomness is represented by a deterministic uncertainty set containing worst-case scenarios. The objective is minimizing the worst-case costs regarding all possible realizations of the uncertainty set. The two-stage adaptive optimization framework has been the subject of many contributions [29, 30, 32]. One difficulty appears in this framework is that the robust counterpart of the second-stage problem is bilinear and non-convex. This bilinear problem can be either solved by the outer approximation method [30] (only local optimality is guaranteed) or rewritten into a mixed integer linear program (MILP) using the big-M method [29, 32] (under some assumptions on the uncertainty set [32]) and solved by generic MILP solvers. Although RO provides security against the worst-case scenario, it may also yield over-conservative solutions resulting from the sheer ignorance of underlying probabilistic information. Paper [34] proposes a UC formulation combining SP and RO to melt the bright sides of both approaches.

Note that the SP assumes the underlying probability distribution of uncertainties to be

precisely known, whereas RO ignores the probabilistic information. In practice, the probability distribution truly exists but must be estimated from historical data and is, therefore, itself uncertain. To better modeling and tackling uncertainties, in several very recent studies, distributionally robust optimization (DRO) has been introduced to power system optimization problems including unit commitment [35, 36], energy reverse dispatch [37, 38], reverse scheduling [39, 40] and DC optimal power flow [41, 42]. DRO assumes that the true probability distribution of uncertain parameters lies in an ambiguity set (of probability distributions) and immunizes the operation strategies against all distributions in the ambiguity set. Different ways to construct the ambiguity set leads to different DRO approaches with the different degree of conservatism and computational efficiency. Paper [35] adopted a scenario-based approach where the random variable representing wind generation is assumed to have finite support. Statistical inference technique is employed to construct ambiguity sets for the discrete distributions. This method is data-driven and more data leads to the less conservative solution. Other papers assume that the random variables have continuous distributions. In [36], the support of a one-dimensional random variable is partitioned into several segments, and the ambiguity set imposes an upper bound for the expectation in each segment. As the number of segments increases, the probability distribution can be characterized with more details. The ambiguity sets employed in [37–40, 42] are the sets of all probability distributions with given mean and covariance. Paper [38] further assumes unimodality to reduce conservatism, and reference [42] further considers the uncertainties of mean and covariance. Different from all others, paper [41, 43] assumes the distribution type is known as a priori while the mean and covariance are subject to uncertainties. To clear the jungle of ambiguity sets, we raise four criteria to judge the quality of ambiguity sets.

1. *Tractability:* DRO problems are reformulated as deterministic optimization problems to be solved by numerical methods. The numerical tractability of the corresponding deterministic problems decides the solvability and practicability of the DRO approach. Therefore, the designed ambiguity set must allow a tractable and efficient reformulation of the DRO problems. For example, the DRO problems in [35, 36] are reformulated as (mixed integer) linear programmings (LP) whereas those in [37–40, 42] are reformulated as semidefinite programmings (SDP). LP is much more tractable than SDP.
2. *Statistical Foundation:* Is there a sound statistical foundation for determining the parameters of the ambiguity set? For instance, the ambiguity set in [35] is based on the statistical inference which guarantees that the ambiguity set contains the true distribution with given confidence level. In contrast, no theory is provided in [36] to guide the parameter selection.
3. *Scalability:* How does the computational burden change when encoding more detailed probabilistic information in the ambiguity set? For example, the numbers of decision

variables and constraints of the deterministic reformulations grow linearly with the number of bins in [35] and the number of segments in [36].

4. *Data-exploiting Ability:* Can the ambiguity set become smaller when more historical data is available? Intuitively, the more data we have, the more accurate we can deduce about the underlying true distribution, which leads to smaller ambiguity set. DRO with smaller ambiguity set results in the less conservative solution. Such data-exploiting ability is evident in the ambiguity set in [35].

This chapter discusses a data-driven affinely adjustable distributionally robust unit commitment (AA-DRUC). The contribution is twofold. Firstly, we propose a novel ambiguity set based on a non-parametric confidence band of the cumulative distribution function (CDF) of the random variable. Secondly, assuming the generation units respond affinely to the total forecasting error of renewable generation, we present a UC problem formulation that minimizes the expected operation costs corresponding to the worst-case distribution in the proposed ambiguity set while explicitly considers spinning reserve, wind curtailment, and load shedding. Distributionally robust chance constraints are employed to guarantee reserve and transmission adequacy. The proposed method possesses the following features:

1. The problem is finally formulated as a mixed integer linear programming (MILP) for which off-the-shelf solvers are available. (tractability)
2. All parameters in the ambiguity set are automatically obtained through non-parametric inference. (statistical foundation)
3. The method is data-driven in the sense that no prior knowledge about the probability distribution of the uncertainties is needed and the historical data is directly incorporated in the solution process. The more historical data is available, the less conservative the solution is. (data-exploiting ability)
4. The scale of the MILP remains unchanged as the amount of historical data and the number of uncertain renewable sources increase. (scalability)

The rest of this chapter is organized as follows. Section 2.2 introduces the ambiguity set constructed from confidence band of CDF. Section 2.3 presents the problem formulation and solution approach of the AA-DRUC. Case studies are reported in section 2.4. Finally, we draw the conclusion and make discussion in section 2.5.

Nomenclature

$\mathcal{B}, \mathcal{L}, \mathcal{T}$	Set of all buses, lines and time periods.
$\mathcal{G}_b$	Set of all generators at bus b.
$s_i^{b,up}/s_i^{b,dn}$	Start-up/shut-down costs of unit i at bus b.
$F_i^b(\cdot)$	Cost function of unit i at bus b.
$d_i^{b,up}/d_i^{b,dn}$	Upward/downward reserve availability price of unit i at bus b.

$f_i^{b,up}/f_i^{b,dn}$ Upward/downward reserve procurement price of unit i at bus b.
C_{ls}/C_{wc} Penalty price of load shedding / wind curtailment.
$M_i^{b,up}/M_i^{b,dn}$ Minimum up/down time of unit i at bus b.
$R_i^{b,up}/R_i^{b,dn}$ Ramp up/down rate limit of unit i at bus b.
$\overline{R}_i^{b,up}/\overline{R}_i^{b,dn}$ Start-up ramp-up / shut-down ramp-down rate limit of unit i at bus b.
L_i^b/U_i^b Output power lower/upper bound of unit i at bus b.
C_l Capacity of transmission line l.
π_l^b Load shift factor from bus b to line l.
w_{it}^b Binary decision variable: "1" if unit i at bus b is on in time t; "0" otherwise.
u_{it}^b Binary decision variable: "1" if unit i at bus b is started up in time t; "0" otherwise.
v_{it}^b Binary decision variable: "1" if unit i at bus b is shut down in time t; "0" otherwise.
x_{it}^b AGC setting point of unit i at bus b in time t.
a_{it}^b AGC participation factor of unit i at bus b in time t.
$r_{it}^{b,up}$ Upward reserve of unit i at bus b in time t.
$r_{it}^{b,dn}$ Downward reserve of unit i at bus b in time t.
$\tilde{p}_{it}^b$ Uncertain actual power output of unit i at bus b in time t.
$\tilde{l}_t^b$ Uncertain composite load at bus b in time t.
$\hat{l}_t^b$ Forecasted composite load at bus b in time t.
$\mathbb{P}$ A probability measure/distribution.
$\mathbb{E}_{\mathbb{P}}$ Expectation respect to probability measure $\mathbb{P}$.
$\mathcal{P}_0(\Xi)$ Set of all probability measures with support Ξ.
$I_{\mathcal{S}}(\cdot)$ Indicator function of set Ξ, i.e. $I_{\Xi}(x) = 1$ when $x \in \Xi$ and 0 otherwise.
$(x)^+$ $\max\{x, 0\}$.

2.2 Ambiguity Set for Univariate Distribution

Consider a one-dimensional random variable ξ whose probability distribution is unknown whereas sample set S is available. The ambiguity set $\mathcal{P}$ is a set of probability distributions consistent with observed sample set S. $\mathcal{P}$ should have the following desired properties: 1) $\mathcal{P}$ contains the underlying true probability distribution; 2) to reduce conservatism, $\mathcal{P}$ can be made as small as possible by incorporating more observed data; 3) the structure of $\mathcal{P}$ allows the reformulation of distributionally robust optimization problems into tractable deterministic problems. In this section, we provide one approach to construct the ambiguity set $\mathcal{P}$ with above properties based on confidence bands for cumulative distribution function (CDF) from non-parametric statistics.

Let $F(x) = \mathbb{P}^*\{\xi \leq x\}$ be the CDF of true distribution $\mathbb{P}^*$. The $1-\alpha$ confidence bands

for $F(x)$ is a pair of sample-dependent functions $\underline{P}(x)$ and $\overline{P}(x)$ for which $\underline{P}(x) \leq F(x) \leq \overline{P}(x),\ \forall x \in \mathbb{R}$ with probability $1-\alpha$ over the choice of sample set S. Deriving finite sample confidence bands of CDF is a basic problem in non-parametric statistics [70]. The most widely used method is based on Kolmogorov-Smirnov (KS) statistic [70], but the bands obtained by KS test are well-known to be unfavorably wide in the tails of the distribution. Recently, the Dirichlet method provides even sensitivity in different parts of the distribution and the closed-form approximation formula given in [71] make this method commendably easy to use. We summarize the method from [71] in the following Lemma.

Lemma 2.1 ([71]): Let $S = \{\hat{\xi}^{(1)}, \hat{\xi}^{(2)}, \cdots, \hat{\xi}^{(n)}\}$ be the ascendingly ordered sample set of random variable ξ generated independently according to true distribution $\mathbb{P}^*$ with continuous CDF $F(x)$. $B^{\alpha}_{k,n}$ denotes the α-quantile of the $\beta(k, n+1-k)$ distribution. For given n and α, define $\underline{p}_k = B^{\tilde{\alpha}/2}_{k,n}$ and $\overline{p}_k = B^{1-\tilde{\alpha}/2}_{k,n}$ where

$$\tilde{\alpha} = \exp\left(-c_1(\alpha) - c_2(\alpha)\sqrt{\ln[\ln(n)]} - c_3(\alpha)[\ln(n)]^{c_4(\alpha)}\right) \tag{2-1}$$

with $c_1(\alpha) = -2.75 - 1.04\ln(\alpha)$, $c_2(\alpha) = 4.76 - 1.20\alpha$, $c_3(\alpha) = 1.15 - 2.39\alpha$, and $c_4(\alpha) = -3.96 + 1.72\alpha^{0.171}$. Add $\hat{\xi}^{(0)} = -\infty$ and $\hat{\xi}^{(n+1)} = \infty$ to the ascending sequence of the sample set S, and define $\underline{p}_0 = 0$ and $\overline{p}_{n+1} = 1$. Then

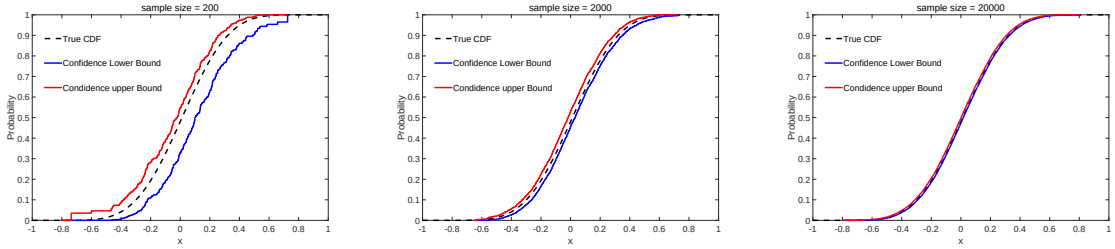
$$\underline{P}(x) = \max\{\underline{p}_k : \hat{\xi}^{(k)} \leq x\} \tag{2-2}$$

$$\overline{P}(x) = \min\{\overline{p}_k : \hat{\xi}^{(k)} \leq x\} \tag{2-3}$$

are the $1-\alpha$ confidence bands for $F(x)$. □

The underlying principle for Lemma 2.1 is that the spacings $F(\hat{\xi}^{(1)})$, $F(\hat{\xi}^{(2)}) - F(\hat{\xi}^{(1)})$,$\cdots$, $F(\hat{\xi}^{(n)}) - F(\hat{\xi}^{(n-1)})$ are random variable having n-variate Dirichlet distribution $D(1, \cdots, 1; 1)$ if S is independently sampled from an identical continuous distribution. Equivalently, the random variables $F(\hat{\xi}^{(1)})$, $F(\hat{\xi}^{(2)})$, $\cdots$, $F(\hat{\xi}^{(n)})$ follow the ordered n-variate Dirichlet distribution $D^*(1, \cdots, 1; 1)$. Then marginal distributions are $F(\hat{\xi}^{(k)}) \sim \beta(k, n+1-k)$, for $k = 1, \cdots, n$. Finally, the ordered Dirichlet distribution $D^*(1, \cdots, 1; 1)$ determines the pointwise $1-\tilde{\alpha}(\alpha, n)$ coverage of $F(\hat{\xi}^{(k)})$ that yields the overall $1-\alpha$ coverage of the CDF. Goldman and Kaplan gives the close-form approximation formula (2-1) for the mapping $\tilde{\alpha}(\alpha, n)$ through extensive simulations [71].

Therefore, given sample set S and significance level $1-\alpha$, the confidence bands of CDF can be readily computed using Lemma 2.1. α is generally set to be equal to or smaller than 0.1 in engineering practice, and it is set to be 0.05 in this chapter. Note that $\underline{P}(x)$ and $\overline{P}(x)$ have the following properties: 1) they are stair-step functions that take values $\underline{p}_k$ and $\overline{p}_k$ at $\hat{\xi}^{(k)}$, respectively; 2) the empirical CDF $\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{\hat{\xi}^{(i)} \leq x\}}$ is lower and upper bounded



(a) 200 Samples (b) 2000 Samples (c) 20000 Samples

Figure 2-1 Confidence bands of CDF constructed from 200, 2000 and 20000 samples of the random variable.

by $\underline{P}(x)$ and $\overline{P}(x)$, i.e. $\underline{P}(x) \leq \hat{F}(x) \leq \overline{P}(x)$; 3) as the size of the sample set $n \to \infty$, $\sup|\underline{P}(x) - \overline{P}(x)| \to 0$. In other words, $\underline{P}(x)$ and $\overline{P}(x)$ represent the reliable information that can be extracted from finite samples and the information becomes more and more accurate as the size of the sample set grows. Fig. 2-1 illustrates the evolution of the obtained CDF confidence bands as the size of the sample set increases.

Note that confidence bands for CDF do not contain the support information of the random variable. We further introduce Devroye-Wise method [72] to estimate the support $[\underline{\xi}, \overline{\xi}]$. Define $\delta = \max_{1\leq i\leq n} |\hat{\xi}^{(i+1)} - \hat{\xi}^{(i)}|$. It is suggested in [73] to choose $[\underline{\xi}, \overline{\xi}] = [\hat{\xi}^{(1)} - \delta/2, \hat{\xi}^{(n)} + \delta/2]$ which is proved to converge to the true support in probability.

Based on the confidence bands for CDF and the estimated support, the ambiguity set $\mathcal{P}$ employed in this chapter takes the form

$$\mathcal{P} = \left\{ \mathbb{P} \in \mathcal{P}_0([\underline{\xi}, \overline{\xi}]) \middle| \mathbb{P}\{\xi \leq \hat{\xi}^{(k)}\} \in [\underline{p}_k, \overline{p}_k],\ k = 1, \cdots, n \right\} \tag{2-4}$$

where $\mathcal{P}_0([\underline{\xi}, \overline{\xi}])$ denotes the set of all probability measures whose supports are the interval $[\underline{\xi}, \overline{\xi}]$. The proposed structure of ambiguity $\mathcal{P}$ is designed to encode the information from confidence bands for CDF and does not assume any prior knowledge about the distribution type. Due to the convergence property of the confidence bands shown in Fig. 2-1, the ambiguity set $\mathcal{P}$ is made smaller and smaller by incorporating more and more historical data. Moreover, the structure defined in (2-4) allows very efficient reformulation of distrituionally robust optimization problems, which will be analyzed in section 2.3. Compared with uncertainty set used in RO, ambiguity set is a set of probability distributions (measures) while uncertainty set is a set of possible realizations of the random variable. Analogous to RO which considers the worst-case realization in the uncertainty set, the proposed DRO considers the worst-case distribution in the ambiguity set.

2.3 Affinely Adjustable Distributionally Robust Unit Commitment

2.3.1 Feasible Set

The constraints for the binary commitment variables are as follows: $\forall i \in \mathcal{G}_b, b \in \mathcal{B}, t \in \mathcal{T}$

$$
\begin{aligned}
&- w^b_{i(t-1)} + w^b_{it} - w^b_{ik} \leq 0, \\
&\qquad \forall k \in [t+1, \min\{t + M^{b,up}_i - 1, T\}] && \text{(2-5)} \\
&w^b_{i(t-1)} - w^b_{it} + w^b_{ik} \leq 1, \\
&\qquad \forall k \in [t+1, \min\{t + M^{b,dn}_i - 1, T\}] && \text{(2-6)} \\
&- w^b_{i(t-1)} + w^b_{it} - u^b_{it} \leq 0 && \text{(2-7)} \\
&w^b_{i(t-1)} - w^b_{it} - v^b_{it} \leq 0 && \text{(2-8)} \\
&w^b_{it}, u^b_{it}, v^b_{it} \in \{0, 1\} && \text{(2-9)}
\end{aligned}
$$

where constraints (2-5) and (2-6) represent the minimum up-time and minimum down-time restrictions, and constraints (2-7) and (2-8) indicate the relationship between on/off status and start up/down operations.

As for the continuous dispatch problem, we adopt the affinely adjustable approach [43, 74–77], i.e. the actual power outputs of generators respond affinely to the total forecasting error of composite loads:

$$\tilde{p}^b_{it} = x^b_{it} + a^b_{it} \sum_{b \in \mathcal{B}} (\tilde{l}^b_t - \hat{l}^b_t) \tag{2-10}$$

where x^b_{it} is the generator setting point and a^b_{it} is the generator participation factor in response to the total difference between uncertain composite load $\tilde{l}^b_t$ and its forecast value $\hat{l}^b_t$. This affine policy has its own limitations. First, the affine policy is only a conservative approximation to the optimal recourse action. Second, only system-level probabilistic information is exploited. However, it possesses some indispensable advantages as follows. 1) Tractability: the robust counterparts of affinely adjustable approach is usually tractable convex problems whereas those of fully adjustable approach are non-convex problems. 2) Scalability: due to such aggregated treatment with uncertainties, we only need to consider a one-dimensional rather than high-dimensional random variable in each time period. Therefore, the computation burden does not increase with the number of uncertain sources; moreover, thanks to the designed ambiguity set and the related reformulation technique discussed later, the computation burden remains unchanged when using more historical data. 3) Practicability: due to the system operator's ability to aggregate uncertainty across all renewable sources, the system-wide forecast is much more accurate. System-wide wind and load prediction is usually used for day-ahead generation scheduling in practice. In addition, affinely adjustable approach is directly compatible with

automatic generation control (AGC) systems where generators respond to area control error (ACE) according to participation factors [61].

To reduce the dimensions of random variables, define

$$\tilde{s}_t = \sum_{b\in\mathcal{B}} (\tilde{l}_t^b - \hat{l}_t^b) \tag{2-11}$$

$$\tilde{h}_t^l = \sum_{b\in\mathcal{B}} \pi_l^b (\tilde{l}_t^b - \hat{l}_t^b) \tag{2-12}$$

Let $\mathbb{P}_t^l$ denote the probability distribution for 2-dimensional random variable $(\tilde{s}_t, \tilde{h}_t^l)$ with marginal distributions $\mathbb{P}_t^s$ and $\mathbb{P}_t^h$. Using the historical data of $\tilde{s}_t$ and $\tilde{h}_t^l$, we can construct the ambiguity sets as defined in (2-4) for $\mathbb{P}_t^s$ and $\mathbb{P}_t^h$, denoted as $\mathcal{P}_t^s$ and $\mathcal{P}_t^h$, respectively. Then the ambiguity set of joint distribution $\mathbb{P}_t^l$ is defined as $\mathcal{P}_t^l = \left\{ \mathbb{P}_t^l \in \mathcal{P}_0(\mathcal{R}^2) \middle| \mathbb{P}_t^s \in \mathcal{P}_t^s, \mathbb{P}_t^h \in \mathcal{P}_t^h \right\}$.

The feasible set of the economic dispatch is described as follows: $\forall i \in \mathcal{G}_b, \forall b \in \mathcal{B}, \forall l \in \mathcal{L}, \forall t \in \mathcal{T}$

$$\sum_{b\in\mathcal{B}} \sum_{i\in\mathcal{G}_b} x_{it}^b - \sum_{b\in\mathcal{B}} \hat{l}_t^b = 0 \tag{2-13}$$

$$\sum_{b\in\mathcal{B}} \sum_{i\in\mathcal{G}_b} a_{it}^b = 1 \tag{2-14}$$

$$0 \le a_{it}^b \le w_{it}^b \tag{2-15}$$

$$L_i^b w_{it}^b + r_{it}^{b,dn} \le x_{it}^b \le U_i^b w_{it}^b - r_{it}^{b,up} \tag{2-16}$$

$$(x_{it}^b + r_{it}^{b,up}) - (x_{i(t-1)}^b - r_{i(t-1)}^{b,dn}) \le (2 - w_{i(t-1)}^b - w_{it}^b)\overline{R}_i^{b,up} + (1 + w_{i(t-1)}^b - w_{it}^b) R_i^{b,up} \tag{2-17}$$

$$(x_{i(t-1)}^b + r_{i(t-1)}^{b,up}) - (x_{it}^b - r_{it}^{b,dn}) \le (2 - w_{i(t-1)}^b - w_{it}^b)\overline{R}_i^{b,dn} + (1 - w_{i(t-1)}^b + w_{it}^b) R_i^{b,dn} \tag{2-18}$$

$$\inf_{\mathbb{P}_t^s \in \mathcal{P}_t^s} \mathbb{P}_t^s \{ -r_{it}^{b,dn} \le a_{it}^b \tilde{s}_t \le r_{it}^{b,up}, \forall i \in \mathcal{G}_b, b \in \mathcal{B} \} \ge 1 - \beta \tag{2-19}$$

$$\inf_{\mathbb{P}_t^l \in \mathcal{P}_t^l} \mathbb{P}_t^l \left\{ -C_l \le \sum_{b\in\mathcal{B}} \pi_l^b \left(\sum_{i\in\mathcal{G}_b} (x_{it}^b + a_{it}^b \tilde{s}_t) - \hat{l}_t^b \right) - \tilde{h}_t^l \le C_l \right\} \ge 1 - \beta - \gamma \tag{2-20}$$

Equality constraints (2-13) and (2-14) together ensure the total generation-consumption balance at every time period in the presence of forecasting errors of composite loads. Constraint (2-15) enforces limits on generator participation factors. Constraints (2-16), (2-17) and (2-18) together ensure upward and downward reserve of each generator are actually procurable considering generator capacity and ramp rate limits. Distributionally robust chance constraints (DRCC) (2-19) and (2-20) guarantee the adequacy of the generator reserve and transmission line capacity with high probability. DRCC ensures the system reliability for all probability distributions in the ambiguity sets, which provide robustness for the system operation.

The DRCC (2-19) and (2-20) admit deterministic safe approximation in light of the defi-

nition of the ambiguity set (2-4), which is revealed in the following two lemmas.

Lemma 2.2： Let $\underline{S}_t(x)$ and $\overline{S}_t(x)$ be the confidence bands for the CDF of random variable $\tilde{s}_t$ defined in (2-2) and (2-3). If $-r_{it}^{b,dn} \le a_{it}^b \tilde{s}_t \le r_{it}^{b,up}$, $\forall i \in \mathcal{G}_b, b \in \mathcal{B}$, $\forall \tilde{s}_t \in [\underline{s}'_t, \overline{s}'_t]$ where $\underline{s}'_t = \overline{S}_t^{-1}(\beta_1)$ and $\overline{s}'_t = \underline{S}_t^{-1}(1-\beta_2)$ with $\beta_1 + \beta_2 = \beta$, then DRCC (2-19) is satisfied.□

Proof： Just need to notice that $\forall \mathbb{P}_t^s \in \mathcal{P}_t^s$,

$$\begin{aligned} &\mathbb{P}_t^s\{\tilde{s}_t \notin [\underline{s}'_t, \overline{s}'_t]\} \\ =&\mathbb{P}_t^s\{\tilde{s}_t < \underline{s}'_t\} + \mathbb{P}_t^s\{\tilde{s}_t > \overline{s}'_t\} \\ \le&\overline{S}_t(\underline{s}'_t) + 1 - \underline{S}_t(\overline{s}'_t) = \beta. \end{aligned} \tag{2-21}$$

■

Therefore, DRCC (2-19) can be safely replaced by the deterministic robust counterpart: $\forall i \in \mathcal{G}_b, b \in \mathcal{B}$

$$-r_{it}^{b,dn} \le a_{it}^b \underline{s}'_t \tag{2-22a}$$

$$a_{it}^b \overline{s}'_t \le r_{it}^{b,up}. \tag{2-22b}$$

Lemma 2.3： Let $\underline{S}_t(x)$ and $\overline{S}_t(x)$ be the confidence bands for the CDF of random variable $\tilde{s}_t$, and $\underline{H}_t^l(x)$ and $\overline{H}_t^l(x)$ be the confidence bands for the CDF of random variable $\tilde{h}_t^l$. If $-C_l \le \sum_{b\in\mathcal{B}} \pi_l^b \left(\sum_{i\in\mathcal{G}_b}(x_{it}^b + a_{it}^b \tilde{s}_t) - \hat{l}_t^b\right) - \tilde{h}_t^l \le C_l$, $\forall \tilde{s}_t \in [\underline{s}'_t, \overline{s}'_t]$, $\tilde{h}_t^l \in [\underline{h}'_{tl}, \overline{h}'_{tl}]$, $\forall l \in \mathcal{L}$ where $\underline{s}'_t = \overline{S}_t^{-1}(\beta_1)$, $\overline{s}'_t = \underline{S}_t^{-1}(1-\beta_2)$ with $\beta_1 + \beta_2 = \beta$, and $\underline{h}'_{tl} = (\overline{H}_t^l)^{-1}(\gamma/2)$ and $\overline{h}'_{tl} = (\underline{H}_t^l)^{-1}(1-\gamma/2)$, then DRCC (2-20) is satisfied. □

Proof： Just need to notice that $\forall \mathbb{P}_t^l \in \mathcal{P}_t^l$

$$\begin{aligned} &\mathbb{P}_t\left\{\tilde{s}_t \notin [\underline{s}'_t, \overline{s}'_t] \text{ or } \tilde{h}_t^l \notin [\underline{h}'_{tl}, \overline{h}'_{tl}]\right\} \\ \le&\mathbb{P}_t\{\tilde{s}_t < \underline{s}'_t\} + \mathbb{P}_t\{\tilde{s}_t > \overline{s}'_t\} + \mathbb{P}_t\{\tilde{h}_t^l < \underline{h}'_{tl}\} + \mathbb{P}_t\{\tilde{h}_t^l > \overline{h}'_{tl}\} \\ =&\mathbb{P}_t^s\{\tilde{s}_t < \underline{s}'_t\} + \mathbb{P}_t^s\{\tilde{s}_t > \overline{s}'_t\} + \mathbb{P}_t^l\{\tilde{h}_t^l < \underline{h}'_{tl}\} + \mathbb{P}_t^l\{\tilde{h}_t^l > \overline{h}'_{tl}\} \\ \le&\overline{S}_t(\underline{s}'_t) + 1 - \underline{S}_t(\overline{s}'_t) + \overline{H}_t^l(\underline{h}'_{tl}) + 1 - \underline{H}_t^l(\overline{h}'_{tl}) = \beta + \gamma \end{aligned} \tag{2-23}$$

■

Similarly, DRCC (2-20) is replaced by: $\forall l \in \mathcal{L}$

$$\sum_{b\in\mathcal{B}} \pi_l^b \left(\sum_{i\in\mathcal{G}_b}(x_{it}^b + a_{it}^b \underline{s}'_t) - \hat{l}_t^b\right) - \underline{h}'_{tl} \le C_l \tag{2-24a}$$

$$-C_l \le \sum_{b\in\mathcal{B}} \pi_l^b \left(\sum_{i\in\mathcal{G}_b}(x_{it}^b + a_{it}^b \underline{s}'_t) - \hat{l}_t^b\right) - \overline{h}'_{tl} \tag{2-24b}$$

$$\sum_{b\in\mathcal{B}} \pi_l^b \left(\sum_{i\in\mathcal{G}_b}(x_{it}^b + a_{it}^b \overline{s}'_t) - \hat{l}_t^b\right) - \underline{h}'_{tl} \le C_l \tag{2-24c}$$

$$-C_l \le \sum_{b\in\mathcal{B}} \pi_l^b \left(\sum_{i\in\mathcal{G}_b}(x_{it}^b + a_{it}^b \overline{s}'_t) - \hat{l}_t^b\right) - \overline{h}'_{tl} \tag{2-24d}$$

Remark 2.1： In Lemma 2.2 and Lemma 2.3, any positive values of β_1 and β_2 with $\beta_1 + \beta_2 = \beta$ will make (2-22)(2-24) safe approximations to (2-19)(2-20). Here β_1 and β_2 have clear engineering meanings. β_1 is the tolerable probability of wind curtailment whereas β_2 is the tolerable probability of load shedding. We leave them as tuning parameters of the method and let the users choose suitable values according to the system reliability standards. □

Remark 2.2： Note that the ambiguity set for $\mathbb{P}_t^l$ is formed by directly combining the information from its marginal distributions. In other words, only marginal distributional information is encoded in $\mathcal{P}_t^l$ and the dependency information of the two marginals are not exploited. This inevitably brings additional conservatism to the chance constraints for the transmission line flow. However, this treatment avoids the consideration of high-dimensional statistics, which contributes to the highly scalable algorithm. □

2.3.2 Objective Function

By Lemma 2.2 and Lemma 2.3, the system can safely respond to the total forecasting error $\tilde{s}_t$ in the interval $[\underline{s}'_t, \overline{s}'_t]$. To ensure the system security, the system operator resorts to load shedding when $\tilde{s}_t$ exceeds $\overline{s}'_t$ and initiates wind curtailment when $\tilde{s}_t$ goes below $\underline{s}'_t$, shown in Fig. 2-2. Therefore, the total operation costs including those of unit start-up/shut-down,

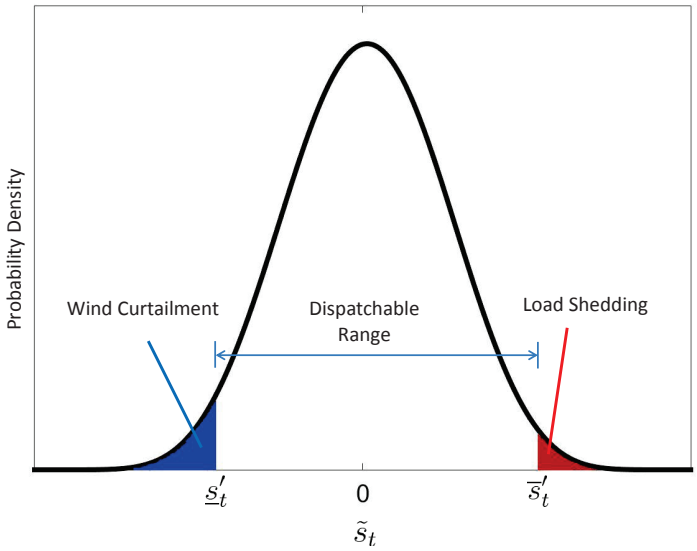


Figure 2-2 Distribution of forecasting error with illustration of wind curtailment and load sheeding.

generation, reserve availability, reserve procurement, load shedding and wind curtailment, are written as

$$\begin{aligned} & F(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{x}, \boldsymbol{a}, \boldsymbol{r}) \\ =& \sum_{t\in\mathcal{T}}\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{G}_b}(s_i^{b,up}u_{it}^b + s_i^{b,dn}v_{it}^b + F_i^b(x_{it}^b) + d_i^{b,up}r_{it}^{b,up} \\ & + d_i^{b,dn}r_{it}^{b,dn}) + \sum_{t\in\mathcal{T}}\max_{\mathbb{P}_t^s\in\mathcal{P}_t^s}\mathbb{E}_{\mathbb{P}_t^s}\{Q_t(\boldsymbol{a}_t, \tilde{s}_t) + C_t(\tilde{s}_t)\} \end{aligned} \tag{2-25}$$

where $\boldsymbol{a}_t = (a_{it}^b)_{i\in\mathcal{G}_b, b\in\mathcal{B}}$, $\boldsymbol{a} = (\boldsymbol{a}_t)_{t\in\mathcal{T}}$ and similar definitions apply for $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{x}$ and $\boldsymbol{r}$; moreover,

$$Q_t(\boldsymbol{a}_t, \tilde{s}_t) = \sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{G}_b}\left(f_i^{b,up}(a_{it}^b\min\{\tilde{s}_t, \overline{s}_t'\})^+ + f_i^{b,dn}(-a_{it}^b\max\{\tilde{s}_t, \underline{s}_t'\})^+\right) \tag{2-26}$$

$$C_t(\tilde{s}_t) = C_{ls}(\tilde{s}_t - \overline{s}_t')^+ + C_{wc}(-\tilde{s}_t + \underline{s}_t')^+. \tag{2-27}$$

Reserve procurement costs is represented by Q_t in which the first and second terms are related to upward and downward reserve procurement, respectively. Also, the first term of C_t denotes load shedding costs and the second represents wind curtailment costs.

Remark 2.3： Note that an explicit formulation for the amount of upward reserve procurement is $(\min\{a_{it}^b\tilde{s}_t, r_{it}^{b,up}\})^+$. We can show $(\min\{a_{it}^b\tilde{s}_t, r_{it}^{b,up}\})^+ = (a_{it}^b\min\{\tilde{s}_t, \overline{s}_t'\})^+$ at the optimal solution of the problem. Suppose $(a_{it}^{b*}, r_{it}^{b,up*})$ is at the optimal solution with $r_{it}^{b,up*} > a_{it}^{b*}\overline{s}_t'$. Observe that $r_{it}^{b,up\#} = a_{it}^{b*}\overline{s}_t'$ is also a feasible solution for the upward reserve (just check constraint (2-16)(2-17)(2-18)(2-22)). In addition, $(\min\{a_{it}^{b*}\tilde{s}_t, r_{it}^{b,up\#}\})^+ = (\min\{a_{it}^{b*}\tilde{s}_t, a_{it}^{b*}\overline{s}_t'\})^+ \leq (\min\{a_{it}^{b*}\tilde{s}_t, r_{it}^{b,up*}\})^+$ and $r_{it}^{b,up\#} < r_{it}^{b,up*}$, so the costs for $(a_{it}^{b*}, r_{it}^{b,up\#})$ is srtictly less than that of $(a_{it}^{b*}, r_{it}^{b,up*})$, which contradicts the optimality of $(a_{it}^{b*}, r_{it}^{b,up*})$. Therefore, we have $r_{it}^{b,up} = a_{it}^b\overline{s}_t'$ at the optimal solution. Then we have $(\min\{a_{it}^b\tilde{s}_t, r_{it}^{b,up}\})^+ = (\min\{a_{it}^b\tilde{s}_t, a_{it}^b\overline{s}_t'\})^+ = (a_{it}^b\min\{\tilde{s}_t, \overline{s}_t'\})^+$ where the second equality follows from the non-negativity of a_{it}^b. It is the same case for the downward reserve. □

Remark 2.4： Note that an explicit formulation for the amount of load shedding should be computed for each bus and then summed up, i.e. $\sum_{b\in\mathcal{B}}(\sum_{i\in\mathcal{G}_b}a_{it}^b\tilde{s}_t - \sum_{i\in\mathcal{G}_b}r_{it}^{b,up})^+$. Following Remark 2.3, we have $\sum_{b\in\mathcal{B}}(\sum_{i\in\mathcal{G}_b}a_{it}^b\tilde{s}_t - \sum_{i\in\mathcal{G}_b}r_{it}^{b,up})^+ = \sum_{b\in\mathcal{B}}(\sum_{i\in\mathcal{G}_b}a_{it}^b\tilde{s}_t - \sum_{i\in\mathcal{G}_b}a_{it}^b\overline{s}_t')^+ = (\tilde{s}_t - \overline{s}_t')^+$ where the second equality comes from non-negativity of a_{it}^b and equation (14). It is also the same case for wind curtailment. □

2.3.3 Evaluation of Worst-case Costs

In the objective function of the proposed formulation, we need to evaluate the worst-case expectation of the piece-wise linear function of the random variable taking the general form

$$\max_{\mathbb{P}_t^s\in\mathcal{P}_t^s}\mathbb{E}_{\mathbb{P}_t^s}\{Q_t(\boldsymbol{a}_t, \tilde{s}_t) + C_t(\tilde{s}_t)\} = \max_{\mathbb{P}_t^s\in\mathcal{P}_t^s}\mathbb{E}_{\mathbb{P}_t^s}\{\max\{f_t^{up}(\tilde{s}_t), f_t^{dn}(\tilde{s}_t)\}\} \tag{2-28}$$

where

$$f_t^{up}(\tilde{s}_t) = g_t^{up}\tilde{s}_t + (C_{ls} - g_t^{up})(\tilde{s}_t - \overline{s}'_t)^+ \tag{2-29}$$

$$f_t^{dn}(\tilde{s}_t) = -g_t^{dn}\tilde{s}_t + (C_{wc} - g_t^{dn})(-\tilde{s}_t + \underline{s}'_t)^+ \tag{2-30}$$

with $g_t^{up} = \sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{G}_b} f_i^{b,up}a_{it}^b$ and $g_t^{dn} = \sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{G}_b} f_i^{b,dn}a_{it}^b$. To show equation (2-28), we only need to be aware of the following equalities:

$$(\min\{\tilde{s}_t, \overline{s}'_t\})^+ = \left(\tilde{s}_t - (\tilde{s}_t - \overline{s}'_t)^+\right)\cdot I_{[0,\infty)}(\tilde{s}_t) \tag{2-31}$$

$$(-\max\{\tilde{s}_t, \underline{s}'_t\})^+ = \left(-\tilde{s}_t - (-\tilde{s}_t + \underline{s}'_t)^+\right)\cdot I_{(-\infty,0]}(\tilde{s}_t) \tag{2-32}$$

which can be verified by a classified calculation and noting that $\overline{s}'_t > 0$ and $\underline{s}'_t < 0$. Then what follows are just steps of direct calculation:

$$\begin{aligned}
&Q_t(\boldsymbol{a}_t, \tilde{s}_t) + C_t(\tilde{s}_t)\\
=&\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{G}_b}\left(f_i^{b,up}(a_{it}^b\min\{\tilde{s}_t, \overline{s}'_t\})^+ + f_i^{b,dn}(-a_{it}^b\max\{\tilde{s}_t, \underline{s}'_t\})^+\right)\\
&\qquad + C_{ls}(\tilde{s}_t - \overline{s}'_t)^+ + C_{wc}(-\tilde{s}_t + \underline{s}'_t)^+\\
=&g_t^{up}(\min\{\tilde{s}_t, \overline{s}'_t\})^+ + g_t^{dn}(-\max\{\tilde{s}_t, \underline{s}'_t\})^+\\
&\qquad + C_{ls}(\tilde{s}_t - \overline{s}'_t)^+ + C_{wc}(-\tilde{s}_t + \underline{s}'_t)^+\\
=&g_t^{up}\left(\tilde{s}_t - (\tilde{s}_t - \overline{s}'_t)^+\right)\cdot I_{[0,\infty)}(\tilde{s}_t) + C_{ls}(\tilde{s}_t - \overline{s}'_t)^+\\
&+ g_t^{dn}\left(-\tilde{s}_t - (-\tilde{s}_t + \underline{s}'_t)^+\right)\cdot I_{(-\infty,0]}(\tilde{s}_t) + C_{wc}(-\tilde{s}_t + \underline{s}'_t)^+\\
=&\left(g_t^{up}\tilde{s}_t + (C_{ls} - g_t^{up})(\tilde{s}_t - \overline{s}'_t)^+\right)\cdot I_{[0,\infty)}(\tilde{s}_t)\\
&\qquad + \left(-g_t^{dn}\tilde{s}_t + (C_{wc} - g_t^{dn})(-\tilde{s}_t + \underline{s}'_t)^+\right)\cdot I_{(-\infty,0]}(\tilde{s}_t)\\
=&f_t^{up}(\tilde{s}_t)\cdot I_{[0,\infty)}(\tilde{s}_t) + f_t^{dn}(\tilde{s}_t)\cdot I_{(-\infty,0]}(\tilde{s}_t)\\
=&\max\{f_t^{up}(\tilde{s}_t), f_t^{dn}(\tilde{s}_t)\}.
\end{aligned}$$

It turns out that the worst-case expectation (2-28) can be evaluated by solving a linear programming (LP). The main results are stated in Lemma 2.4 inspired by [78].

Lemma 2.4： Let $\hat{s}_t^{(1)}, \hat{s}_t^{(2)}, \cdots, \hat{s}_t^{(n)}$ be the ascendingly ordered samples of random variable $\tilde{s}_t$. Without loss of generality, assume $\hat{s}_t^{(k)} \le 0, \forall k \le m$ and $\hat{s}_t^{(k)} > 0, \forall k > m$. $[\underline{s}_t, \overline{s}_t]$ is the estimated support of random variable $\tilde{s}_t$ by Devroye-Wise method. The ambiguity set $\mathcal{P}_t^s$ is constructed as in (2-4), i.e.

$$\mathcal{P}_t^s = \left\{\mathbb{P}\in\mathcal{P}_0([\underline{s}_t, \overline{s}_t])\middle|\mathbb{P}\{\tilde{s}_t \le \hat{s}_t^{(k)}\}\in[\underline{p}_t^k, \overline{p}_t^k], k = 1, \cdots, n\right\} \tag{2-33}$$

For notational convenience, let $\hat{s}_t^{(0)} = \underline{s}_t$, $\hat{s}_t^{(n+1)} = \overline{s}_t$ and $\underline{p}_t^{n+1} = \overline{p}_t^{n+1} = 1$. The worst-case

expectation (2-28) is equal to the optimum of the following LP:

$$\min_{\substack{\underline{\lambda}_t^k,\overline{\lambda}_t^k \\ k=m-1,\cdots,m+2}} \sum_{k=m-1}^{m+2} (\overline{\lambda}_t^k \overline{p}_t^k - \underline{\lambda}_t^k \underline{p}_t^k) + g_t^{dn} s_t^{dn} + g_t^{up} s_t^{up} + C_{wc} m_t^{dn} + C_{ls} m_t^{up}$$
$$\text{s.t.} \begin{cases} \underline{\lambda}_t^k \geq 0,\ \overline{\lambda}_t^k \geq 0, k = m-1, \cdots, m+2 \\ \sum_{i=k}^{m+2} (\overline{\lambda}_t^i - \underline{\lambda}_t^i) + f_t^{up}(\hat{s}_t^{(m+3)}) \geq f_t^{dn}(\hat{s}_t^{(k-1)}), k = m-1, m, m+1 \\ \sum_{i=k}^{m+2} (\overline{\lambda}_t^i - \underline{\lambda}_t^i) + f_t^{up}(\hat{s}_t^{(m+3)}) \geq f_t^{up}(\hat{s}_t^{(k)}), k = m+1, m+2 \end{cases} \tag{2-34}$$

where s_t^{dn}, s_t^{up}, m_t^{dn} and m_t^{up} are defined in (2-35), (2-36), (2-37) and (2-38):

$$s_t^{dn} = \sum_{k=1}^{m-2} [\hat{s}_t^{(k)} + (-\hat{s}_t^{(k)} + \underline{s}_t')^+ - \hat{s}_t^{(k-1)} - (-\hat{s}_t^{(k-1)} + \underline{s}_t')^+] \overline{p}_t^k \tag{2-35}$$

$$s_t^{up} = \sum_{k=m+3}^{n} [\hat{s}_t^{(k)} - (\hat{s}_t^{(k)} - \overline{s}_t')^+ - \hat{s}_t^{(k+1)} + (\hat{s}_t^{(k+1)} - \overline{s}_t')^+] \underline{p}_t^k + [\hat{s}_t^{(n+1)} - (\hat{s}_t^{(n+1)} - \overline{s}_t')^+] \overline{p}_t^{n+1}. \tag{2-36}$$

$$m_t^{dn} = \sum_{k=1}^{m-2} [(-\hat{s}_t^{(k-1)} + \underline{s}_t')^+ - (-\hat{s}_t^{(k)} + \underline{s}_t')^+] \overline{p}_t^k \tag{2-37}$$

$$m_t^{up} = \sum_{k=m+3}^{n} [(\hat{s}_t^{(k)} - \overline{s}_t')^+ - (\hat{s}_t^{(k+1)} - \overline{s}_t')^+] \underline{p}_t^k + (\hat{s}_t^{(n+1)} - \overline{s}_t')^+ \overline{p}_t^{n+1}. \tag{2-38}$$

□

Proof： **Step 1** We first show the worst-case expectation can be evaluated by solving a LP.

The evaluation of worst-case expectation (2-28) is a infinite dimensional linear optimization problem of the form

$$\max_{\mathbb{P} \in \mathcal{P}_0} \int_{[\hat{s}_{(0)}, \hat{s}_{(n+1)}]} \max\{f_t^{up}(\tilde{s}_t), f_t^{dn}(\tilde{s}_t)\} \mathbb{P}(d\tilde{s}_t) \tag{2-39a}$$
$$\text{s.t. } \underline{p}_k \leq \int_{[\hat{s}_{(0)}, \hat{s}_{(n+1)}]} I_{[\hat{s}_{(0)}, \hat{s}_{(k)}]}(\tilde{s}_t) \mathbb{P}(d\tilde{s}_t) \leq \overline{p}_k, \forall k = 1, \cdots, n+1. \tag{2-39b}$$

By using the conic duality theory, assigning dual variables $(\underline{\lambda}_k, \overline{\lambda}_k)$ to constraints (2-39b), the dual of problem (2-39) is given by

$$\inf_{\underline{\lambda}, \overline{\lambda}} \sum_{k=1}^{n+1} (\overline{\lambda}_k \overline{p}_k - \underline{\lambda}_k \underline{p}_k) \tag{2-40a}$$
$$\text{s.t.} \underline{\lambda}_k \geq 0, \overline{\lambda}_k \geq 0,\ 1 \leq k \leq n+1 \tag{2-40b}$$
$$\sum_{k=1}^{n+1} (\overline{\lambda}_k - \underline{\lambda}_k) I_{[\hat{s}_{(0)}, \hat{s}_{(k)}]}(\xi) - \max\{f_t^{up}(\tilde{s}_t), f_t^{dn}(\tilde{s}_t)\} \geq 0, \forall \xi \in [\hat{s}_{(0)}, \hat{s}_{(n+1)}] \tag{2-40c}$$

which is a finite dimensional optimization problem with semi-infinite constraint (2-40c). Strong duality, i.e. the optimal values of (2-39) and (2-40) coincide, is guaranteed by theorem 1 in

[79]. Thus, we focus on the dual problem in the sequel. The semi-infinite constraint (2-40c) is equivalent to

$$\begin{cases} \inf\limits_{\xi\in[\hat{s}_{(0)},\hat{s}_{(n+1)}]} \left\{ \sum\limits_{k=1}^{n+1} (\overline{\lambda}_k - \underline{\lambda}_k) I_{[\hat{s}_{(0)},\hat{s}_{(k)}]}(\xi) - f_t^{up}(\tilde{s}_t) \right\} \geq 0 \\ \inf\limits_{\xi\in[\hat{s}_{(0)},\hat{s}_{(n+1)}]} \left\{ \sum\limits_{k=1}^{n+1} (\overline{\lambda}_k - \underline{\lambda}_k) I_{[\hat{s}_{(0)},\hat{s}_{(k)}]}(\xi) - f_t^{dn}(\tilde{s}_t) \right\} \geq 0 \end{cases} \tag{2-41}$$

Observe that $[\hat{s}_{(0)}, \hat{s}_{(n+1)}]$ can be partitioned into $n+1$ mutually disjoint sets $[\hat{s}_{(k-1)}, \hat{s}_{(k)}]$, $k = 1, \cdots, n+1$. Therefore constraint (2-41) splits into $n+1$ parts, i.e. $\forall k = 1, \cdots, n+1$,

$$\begin{cases} \inf\limits_{\xi\in[\hat{s}_{(k-1)},\hat{s}_{(k)}]} \left\{ \sum\limits_{i=k}^{n+1} (\overline{\lambda}_i - \underline{\lambda}_i) - f_t^{up}(\tilde{s}_t) \right\} \geq 0 \\ \inf\limits_{\xi\in[\hat{s}_{(k-1)},\hat{s}_{(k)}]} \left\{ \sum\limits_{i=k}^{n+1} (\overline{\lambda}_i - \underline{\lambda}_i) - f_t^{dn}(\tilde{s}_t) \right\} \geq 0 \end{cases} \tag{2-42}$$

Since $f_t^{up}(\tilde{s}_t)$ is monotone increasing and $f_t^{dn}(\tilde{s}_t)$ is monotone decreasing, (2-42) is equivalent to

$$\begin{cases} \sum\limits_{i=k}^{n+1} (\overline{\lambda}_i - \underline{\lambda}_i) - f_t^{up}(\hat{s}_{(k)}) \geq 0 \\ \sum\limits_{i=k}^{n+1} (\overline{\lambda}_i - \underline{\lambda}_i) - f_t^{dn}(\hat{s}_{(k-1)}) \geq 0 \end{cases} \tag{2-43}$$

By replacing semi-infinite constraint (2-40c) with constraint (2-43) and noticing $f_t^{up}(\tilde{s}_t)$ and $f_t^{dn}(\tilde{s}_t)$ intersect at $\tilde{s}_t = 0$, problem (2-40) is equivalent to the following LP:

$$\min_{\underline{\lambda}_t, \overline{\lambda}_t} \sum_{k=1}^{n+1} (\overline{\lambda}_t^k \overline{p}_t^k - \underline{\lambda}_t^k \underline{p}_t^k) \tag{2-44}$$

$$\text{s.t.} \begin{cases} \underline{\lambda}_t^k \geq 0,\ \overline{\lambda}_t^k \geq 0, \forall 1 \leq k \leq n+1 \\ \sum\limits_{i=k}^{n+1} (\overline{\lambda}_t^i - \underline{\lambda}_t^i) - f_t^{dn}(\hat{s}_{(k-1)}) \geq 0, \forall 1 \leq k \leq m+1 \\ \sum\limits_{i=k}^{n+1} (\overline{\lambda}_t^i - \underline{\lambda}_t^i) - f_t^{up}(\hat{s}_{(k)}) \geq 0, \forall m+1 \leq k \leq n+1 \end{cases} \tag{2-45}$$

Step 2 We then show the above LP can be significantly simplified. Let

$$x_t^k = \overline{\lambda}_t^k - \underline{\lambda}_t^k,\ 1 \leq k \leq n+1 \tag{2-46}$$

and

$$h_t^k = \begin{cases} f_t^{dn}(\hat{s}_{(k-1)}), & k \leq m \\ \max\{f_t^{dn}(\hat{s}_{(k-1)}), f_t^{up}(\hat{s}_{(k)})\}, & k = m+1 \\ f_t^{up}(\hat{s}_{(k)}), & k \geq m+2. \end{cases} \tag{2-47}$$

LP (2-44)(2-45) is equivalent to

$$\begin{aligned}
&\min_{\boldsymbol{x}_t,\underline{\boldsymbol{\lambda}}_t} \sum_{k=1}^{n+1} \underline{\lambda}_t^k(\overline{p}_t^k - \underline{p}_t^k) + x_t^k\overline{p}_t^k \\
&\text{s.t. } x_t^k + \underline{\lambda}_t^k \geq 0, \underline{\lambda}_t^k \geq 0 \; k = 1, \ldots, n+1 \\
&\quad \sum_{i=k}^{n+1} x_t^i \geq h_t^k.
\end{aligned} \tag{2-48}$$

Since $\overline{p}_t^k - \underline{p}_t^k \geq 0, \forall\, 1 \leq k \leq n+1$, the optimal value for $\underline{\lambda}_t^k$ is always $\max\{-x_t^k, 0\}$, i.e. the LP (2-48) is simplified to

$$\begin{aligned}
&\min_{\boldsymbol{x}_t} \sum_{k=1}^{n+1} \max\{x_t^k\overline{p}_t^k, x_t^k\underline{p}_t^k\} \\
&\text{s.t. } \sum_{i=k}^{n+1} x_t^i \geq h_t^k, \; k = 1, \ldots, n+1.
\end{aligned} \tag{2-49}$$

By Further defining $z_t^k = \sum_{i=k}^{n+1} x_t^i, \; \forall 1 \leq k \leq n+1$ and $z_t^{n+2} = 0$, problem (2-49) can be written as

$$\begin{aligned}
&\min_{\boldsymbol{z}_t} f_t(\boldsymbol{z}_t) = \sum_{k=1}^{n+1} \max\{(z_t^k - z_t^{k+1})\overline{p}_t^k, (z_t^k - z_t^{k+1})\underline{p}_t^k\} \\
&\text{s.t. } z_t^k \geq h_t^k, \; k = 1, \ldots, n+1.
\end{aligned} \tag{2-50}$$

Note that problem (2-50) is convex and $\forall k = 1, \cdots, m-1, m+3, \cdots, n+1$,

$$\begin{aligned}
&\left.\frac{\partial f_t}{\partial z_t^k}\right|_{(h_t^1,\cdots,h_t^{m-1},z_t^m,z_t^{m+1},z_t^{m+2},h_t^{m+3},\cdots,h_t^{n+1})} \\
&= \begin{cases} \overline{p}_t^k - \overline{p}_t^{k-1}, & 1 \leq k \leq m-1 \\ \underline{p}_t^k - \underline{p}_t^{k-1}, & m+3 \leq k \leq n \\ \overline{p}_t^{n+1} - \underline{p}_t^n, & k = n+1 \end{cases} \\
&> 0.
\end{aligned} \tag{2-51}$$

i.e. going along any feasible direction at point $(h_t^1, \cdots, h_t^{m-1}, z_t^m, z_t^{m+1}, z_t^{m+2}, h_t^{m+3}, \cdots, h_t^{n+1})$ lead to the increase of the objective function (more strictly, it is the optimality condition based on tangent cone of the feasible set, see Theorem 3.1 in [80]). Therefore, z_t^k takes optimal value at h_t^k, $\forall k = 1, \cdots, m-1, m+3, \cdots, n+1$. Equivalently, $x_t^k = h_t^k - h_t^{k+1}$, $\forall k = 1, \cdots, m-2, m+3, \cdots, n$ and $x_t^{n+1} = h_t^{n+1}$ at the optimal solution of LP (2-48). It is obvious that $x_t^k > 0, \; 1 \leq k \leq m-2$ and $x_t^k < 0, \; m+3 \leq k \leq n+1$, so $\underline{\lambda}_t^k = 0, \; 1 \leq k \leq m-2$ and $\underline{\lambda}_t^k = -x_t^k, \; m+3 \leq k \leq n+1$. Substituting above results into (2-48) and noticing (2-46), (2-47), (2-29) and (2-30) yield the the simplified LP (2-34), which completes the proof. ■

Note that the LP (2-34) is a small-scale problem with only 4 decision variables and the problem scale is irrelevant to the number of historical data. In other words, incorporating more

historical data does not bring higher computational burden.

Therefore, evaluating the worst-case expectation in the objective function (2-25) with LP (2-34) and replacing the DRCC (2-19)(2-20) with deterministic linear constraints (2-22)(2-24) yield a MILP. To sum up, for given reliability level $1-\beta$ and historical data of forecasting errors, the proposed AA-DRUC takes the form

$$\begin{aligned} O(\beta_1,\beta_2,C_{ls},C_{wc}) =& \min F(\boldsymbol{u},\boldsymbol{v},\boldsymbol{x},\boldsymbol{a},\boldsymbol{r}) \\ & \text{s.t. } (2-5)\sim(2-9),(2-13)\sim(2-18),(2-22)(2-24). \end{aligned} \tag{2-52}$$

2.3.4 Application Modes

In the above problem formulation, we have included both the reliability indices β_1/β_2 and the emergency control prices C_{ls}/C_{wc}. Usually, either of the two is presented as the input to the optimization model in practice. In the conventional vertically integrated systems, electric utilities operate as monopolies and enforce a reliability standard across the network to limit the occurrence of load shedding and wind curtailment. Under such circumstances, reliability indices β_1/β_2 are given as prior knowledge, and the costs of load shedding and wind curtailment are irrelevant to the decision process. We call this application mode the **Mode-I**. To obtain optimal operational strategy in this mode, we only need to solve the problem $O(\beta_1,\beta_2,0,0)$. By comparison, the operation of the modern restructured power systems is price-driven, and the load shedding and wind curtailment are considered ancillary services provided by the consumers and wind farms. Hence the access to these services come at a cost. We name this application mode the **Mode-II**. Note that high reliability results in low emergency control costs but high reverse costs, and vice versa. Therefore, there is a pair of optimal reliability indices β_1^*/β_2^* which minimize the total operation costs, and the solution to the corresponding MILP (2-52) should be considered the optimal operational strategy. To find out the optimal reliability indices β_1^*/β_2^*, we need to solve the higher-level optimization problem $\min_{\beta_1,\beta_2} O(\beta_1,\beta_2,C_{ls},C_{wc})$. Since $O(\beta_1,\beta_2,C_{ls},C_{wc})$ can be conveniently evaluated by solving MILP (2-52), the optimization over β_1,β_2 can be done by any derivative-free search method, e.g. the Nelder-Mead simplex Method [81].

2.3.5 Elimination of Redundant Line Capacity Constraints

In practical operation of power systems, the line capacity constraints (2-24) are only active at very few transmission lines during some periods. Hence most of the constraints in (2-24) are redundant for the optimization problem (2-52). Identifying and eliminating those redundant constraints before solving the MILP could significantly improve computational efficiency. The fast identification method proposed in [82] can be extended to our problem formulation with

minor modification.

Consider the following problems:

$$\begin{aligned}&\Lambda_{t,max}^{l}(\tilde{s}_t,\tilde{h}_t^l)\left(\Lambda_{t,min}^{l}(\tilde{s}_t,\tilde{h}_t^l)\right)\\&=\arg\max_{\boldsymbol{x}_t,\boldsymbol{\alpha}_t}(\min)\sum_{b\in\mathcal{B}}\pi_l^b\left(\sum_{i\in\mathcal{G}_b}(x_{it}^b+a_{it}^b\tilde{s}_t)-\hat{l}_t^b\right)-\tilde{h}_t^l\end{aligned}\tag{2-53}$$

subject to

$$\sum_{b\in\mathcal{B}}\sum_{i\in\mathcal{G}_b}(x_{it}^b+a_{it}^b\tilde{s}_t)=\sum_{b\in\mathcal{B}}\hat{l}_t^b+\tilde{s}_t \tag{2-54a}$$

$$x_{it}^b+a_{it}^b\underline{s}_t'\geq 0 \tag{2-54b}$$

$$x_{it}^b+a_{it}^b\overline{s}_t'\leq U_i^b. \tag{2-54c}$$

where (2-54a) is obtained by multiplying (2-14) by $\tilde{s}_t$ and adding to (2-13); (2-54b)$\sim$(2-54c) are deduced from (2-17), (2-22) by relaxing w_{it}^b to 0 or 1 when necessary. Therefore, the feasible sets of the above optimization problems are relaxations of the feasible set of the original MILP model. Minimization (maximization) w.r.t. the feasible set defined by (2-54) yields a lower (upper) bound of the minimum (maximum) w.r.t the feasible set of the original MILP model. The objective function (2-53) is just the possible line flow at each line in each time period. Similar to the analysis in [82], we have the following lemma.

Lemma 2.5： For any $l\in\mathcal{L}$ and $t\in\mathcal{T}$, we have

- If $\Lambda_{t,max}^{l}(\underline{s}_t,\underline{h}_t^l)\leq C_l$, constraint (2-24a) is inactive;
- If $\Lambda_{t,max}^{l}(\overline{s}_t,\underline{h}_t^l)\leq C_l$, constraint (2-24c) is inactive;
- If $\Lambda_{t,min}^{l}(\underline{s}_t,\overline{h}_t^l)\geq -C_l$, constraint (2-24b) is inactive;
- If $\Lambda_{t,min}^{l}(\overline{s}_t,\overline{h}_t^l)\geq -C_l$, constraint (2-24d) is inactive. □

Define $p_t^b=\sum_{i\in\mathcal{G}_b}p_{it}^b$ where $p_{it}^b=x_{it}^b+a_{it}^b\tilde{s}_t$ and $U^b=\sum_{i\in\mathcal{G}_b}U_i^b$, then we have

$$\begin{aligned}&\Lambda_{t,max}^{l}(\tilde{s}_t,\tilde{h}_t^l)\left(\Lambda_{t,min}^{l}(\tilde{s}_t,\tilde{h}_t^l)\right)\\&=\arg\max_{\boldsymbol{p}_t}(\min)\sum_{b\in\mathcal{B}}\pi_l^b p_t^b-\sum_{b\in\mathcal{B}}\pi_l^b\hat{l}_t^b-\tilde{h}_t^l\\&\text{s.t.}\begin{cases}\sum_{b\in\mathcal{B}}p_t^b=\sum_{b\in\mathcal{B}}\hat{l}_t^b+\tilde{s}_t\\0\leq p_t^b\leq U^b\end{cases}\end{aligned}\tag{2-55}$$

LP (2-55) has a analytical solution according to the analysis in [82]. Let $b_1,b_2,\cdots,b_{|\mathcal{B}|}$ be a permutation of $1,2,\cdots,|\mathcal{B}|$ such that $\{K_l^{b_1},K_l^{b_2},\cdots,K_l^{b_{|\mathcal{B}|}}\}$ are in descending (ascending) order, and there exists an integer $1\leq m\leq|\mathcal{B}|$ such that $\sum_{k=1}^{m-1}U^{b_k}\leq\sum_{b\in\mathcal{B}}\hat{l}_t^b+\tilde{s}_t\leq\sum_{k=1}^{m}U^{b_k}$. Then

$$\begin{aligned}&\Lambda^l_{t,max}(\tilde{s}_t,\tilde{h}^l_t)\left(\Lambda^l_{t,min}(\tilde{s}_t,\tilde{h}^l_t)\right)\\&=\sum_{k=1}^{m-1}(K^{b_k}_l-K^{b_m}_l)U^b+K^{b_m}_l(\sum_{b\in\mathcal{B}}\hat{l}^b_t+\tilde{s}_t)-\sum_{b\in\mathcal{B}}\pi^b_l\hat{l}^b_t-\tilde{h}^l_t.\end{aligned}\tag{2-56}$$

Based on the analytical expression (2-56), Lemma 2.5 gives a computationally cheap way to identify most of the inactive line capacity constraints in (2-24).

2.4 Case Studies

This section presents numerical results on the IEEE 118-bus system and the 1888-bus French very high voltage system. The 118-bus system has 54 units and 186 lines. In the base case, 10 wind farms with each capacity of 80MW are installed across the 118-bus system. The 1888-bus system has 290 units and 2531 lines. In the base case, we install 20 wind farms with each capacity of 500MW over the network. The network data of the test systems is extracted from MATPOWER 5.1 and the unit data is from http://motor.ece.iit.edu/data. The hourly load profile is obtained from [75] and hourly forecasting wind power curves are from NREL WIND Toolkit. In addition, the reserve prices $d^{b,up}_i/d^{b,dn}_i$ and the reserve procurement prices $f^{b,up}_i/f^{b,dn}_i$ are assumed to be 10% and 110% of the coefficients of the linear terms of quadratic generator cost functions. Wind curtailment and load shedding costs are 100\$/MWh and 500\$/MWh, respectively. The proposed method is programmed in MATLAB with Gurobi [83] as the MILP solver running on a Win 8 PC with a 3.0 GHz CPU and 24 GB RAM. The MIP gap is set to be 10e-4.

2.4.1 Distributional Robustness and Data-exploiting Feature

To test the distributional robustness and data-exploiting feature of the proposed method, we generate wind power forecasting errors from four different types of probability distributions, including beta, normal, laplace and hyperbolic [84]. The mean and standard deviation of the forecasting errors are based on the typical day-ahead forecasting errors in U.S. reported in [85], i.e. $\mu = 0.0117$ p.u., $\sigma = 0.1187$ p.u.. Historical sample sets of different sizes are used in the proposed method to reveal the relation between sample size and solution quality. After solving each problem, Monte Carlo simulation (MCS) with another 10^6 samples generated from the corresponding distribution is employed to test the practical and *out-of-sample* performance of the proposed AA-DRUC.

Fig. 2-3 and Fig. 2-4 illustrate the effects of incorporating more historical data on the optimal objective function and the operation costs from MCS on both test systems in application mode-I with $\beta_1 = 0.03, \beta_2 = 0.01$ and $\gamma = 0$. As shown in both figures, no matter what distribution the forecasting errors follow, the optimal values of AA-DRUC objective function always act as the upper bounds of practical operation costs from MCS. This is due to the fact

that the proposed method considers the worst-case distribution in the ambiguity set constructed from historical data, but the underlying true distribution usually differs from the worst one. However, by incorporating more historical data, the ambiguity set shrinks and the worst-cast distribution in the ambiguity set approaches the true distribution. Therefore, as shown in Fig. 2-3 and Fig. 2-4, the values of the optimal objective function and the gap between optimal objective function and operation costs from MCS decrease as the amount of available historical data increases. This reveals the value of data, i.e. the more data we use, the less conservative thus the more economical the operation strategy is. Fig. 2-5∼2-8 further compare the probability of load shedding and wind curtailment from MCS under different types of distributions and different amount of available historical data. It is evident from Fig. 2-5∼2-8 that the prespecified reliability requirements are always satisfied by the proposed method whatever distribution the uncertain forecasting error obeys, which is an explicit demonstration of the distributional robustness of the proposed method. As more and more historical data is available, the practical reliability level of the operation strategy presses on towards the pre-specified reliability indices, which reduces the conservatism of the DRCC. Table 2-1 further lists the detailed unit commitment strategy of 118-bus system when the unkown true distribution is of beta type and 10^4 data is available.

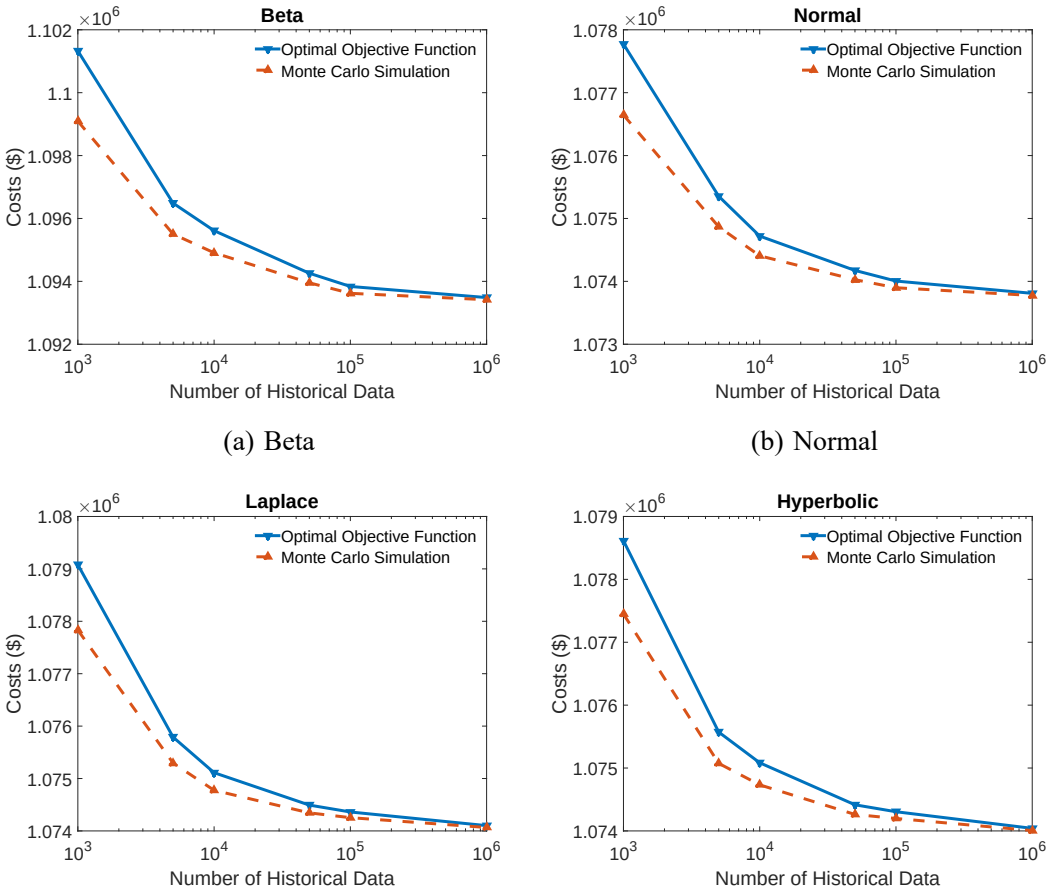
Moreover, Fig. 2-9 and Fig. 2-10 show the evolution of objective function and simulated costs as the increase of available historical data in the application mode-II on both test systems. As β_1, β_2 increases, the dispatchable range illustrated in Fig. 2-2 shrinks and the allocated system reserve decreases reserve costs decreases, whereas the probability and expected costs of wind curtailment and load shedding increase. As a result, there exists a pair of optimal values of β_1, β_2 which can be located by the Nelder-Mead simplex method implemented in MATLAB command `fminsearch`. The optimal reliability indices are labeled near the corresponding optimal values of the objective function on Fig. 2-9 and Fig. 2-10. The results in the application mode-II resemble those of application mode-I: the gap between the optimal objective function and the costs from MCS vanishes as more historical data is available.

2.4.2 Computational Efficiency and Scalability

Table 2-2 presents the solver time in seconds of the proposed method as increasing amount of historical data is available. It confirms the observation we made in section 2.3.3 that the solution time is irrelevant to the number of historical data points employed. Table 2-3, on the other hand, compares the solver time of the proposed method when different numbers of wind farms are installed across the systems. Due to the simple affine policy of reserve procurement employed in the problem formulation, the solver time is also irrelevant to the number of wind farms. Compared with the solver time of the distributionally robust model for UC in [36], the AA-DRUC proposed in this chapter can be solved orders-of-magnitude faster. The high com-

Table 2-1 Unit Commitment Strategy for 118-bus System

unit no. \ time	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
unit 4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
unit 10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 14	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 25	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 29	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 36	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 39	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 40	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 43	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 44	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 45	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
unit 53	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1



(a) Beta (b) Normal

(c) Laplace (d) Hyperbolic

Figure 2-3 Evolution of objective function and simulated costs as the increase of available historical data in application mode-I on 118-bus system. Four types of probability distributions, i.e. beta, normal, laplace and hyperbolic, are used to simulate the underlying true distribution.

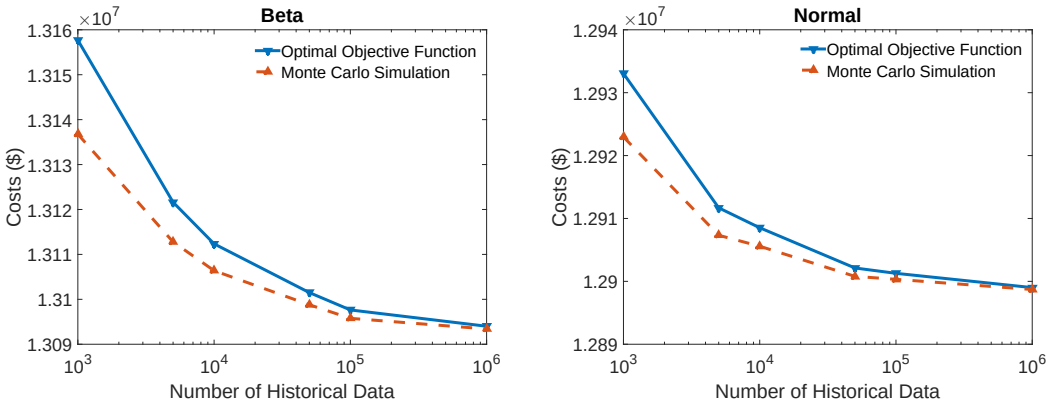
putational efficiency of the proposed method is largely attributed to the redundant constraint identification method discussed in section 2.3.5 which helps to eliminate about $88\% \sim 95\%$ line capacity constraints on both test systems.

Table 2-2 Solver Time (sec.) v.s. Number of Historical Data

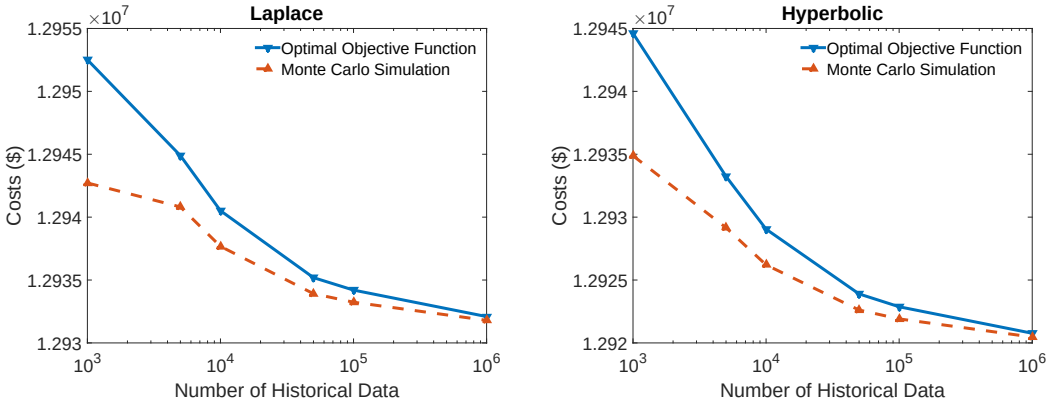
	1.E+03	5.E+03	1.E+04	5.E+04	1.E+05	1.E+06
case118	2.7	4.1	3.7	3.6	3.9	2.8
case1888	42.7	43.6	44.0	41.8	43.5	43.3

2.4.3 Assessing the Conservatism of Affine Policy

Although the affine policy has brought high tractability and scalability for the proposed method, it inevitably introduces some conservatism. We first illustrate how the conservatism



(a) Beta (b) Normal



(c) Laplace (d) Hyperbolic

Figure 2-4 Evolution of objective function and simulated costs as the increase of available historical data in application mode-I on 1888-bus system. Four types of probability distributions, i.e. beta, normal, laplace and hyperbolic, are used to simulate the underlying true distribution.

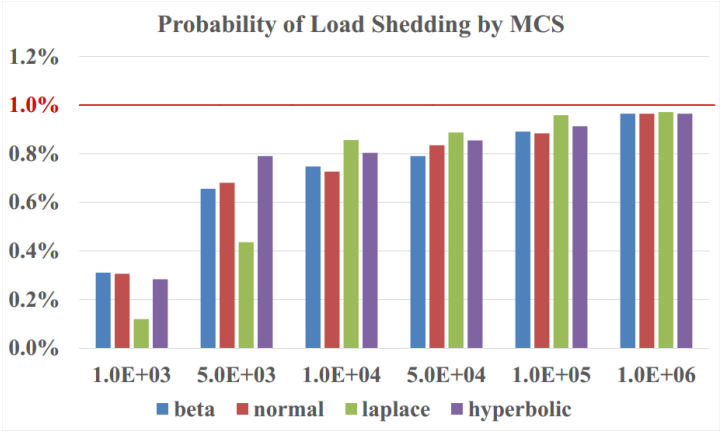


Figure 2-5 Probability of load shedding by MCS on 118-bus system in application mode-I under different number of historical data and different underlying true distributions of uncertainty. The reliability requirement is $< 1.0\%$.

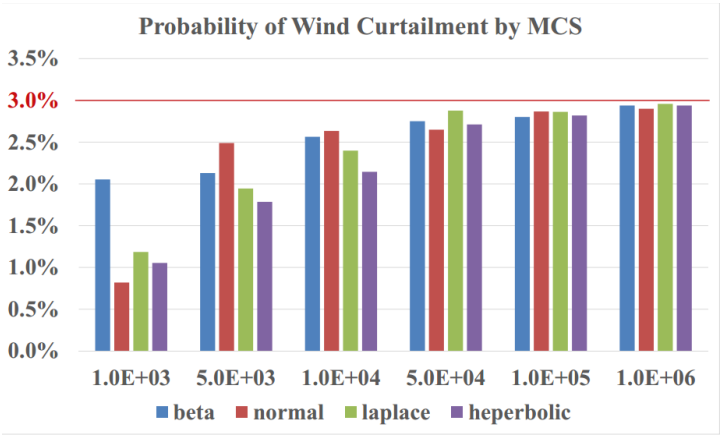


Figure 2-6 Probability of wind curtailment by MCS on 118-bus system in application mode-I under different number of historical data and different underlying true distributions of uncertainty. The reliability requirement is $< 3.0\%$.

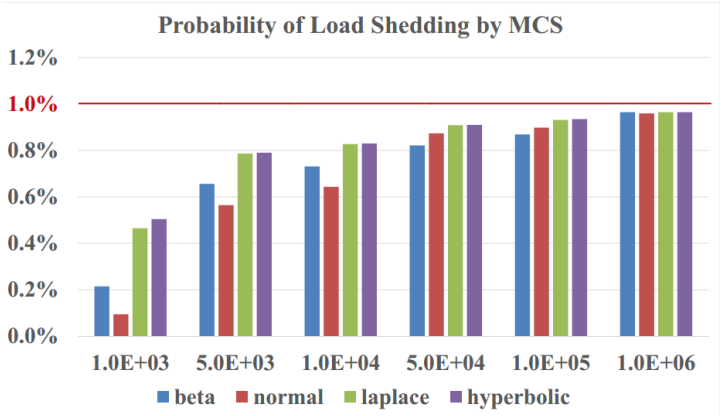


Figure 2-7 Probability of load shedding by MCS on 1888-bus system in application mode-I under different number of historical data and different underlying true distributions of uncertainty. The reliability requirement is $< 1.0\%$.

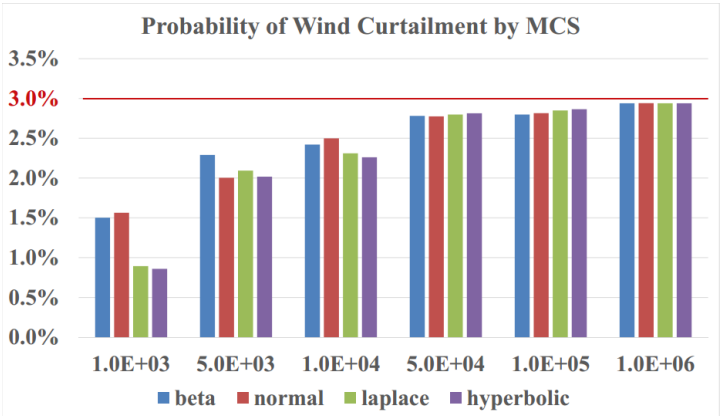


Figure 2-8 Probability of wind curtailment by MCS on 1888-bus system in application mode-I under different number of historical data and different underlying true distributions of uncertainty. The reliability requirement is $< 3.0\%$.

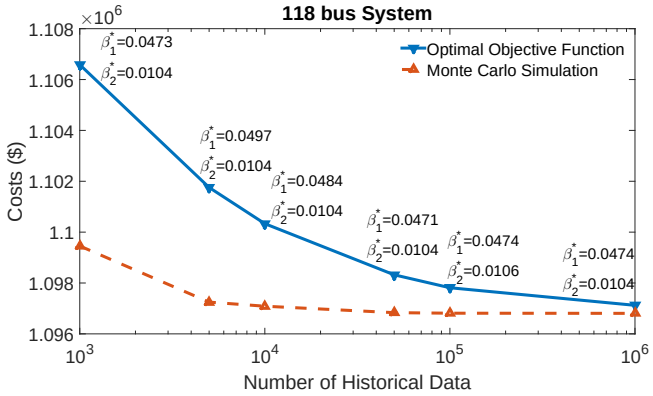


Figure 2-9 Evolution of objective function and simulated costs as the increase of available historical data in application mode-II on 118-bus system. β_1^* and β_2^* are the optimal reliability indices to achieve the corresponding objective value.

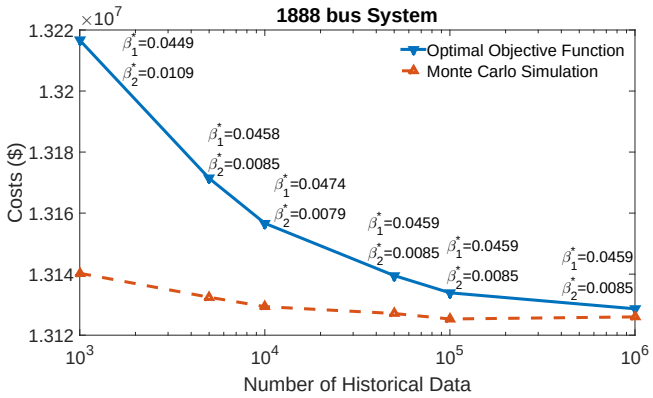


Figure 2-10 Evolution of objective function and simulated costs as the increase of available historical data in application mode-II on 1888-bus system. β_1^* and β_2^* are the optimal reliability indices to achieve the corresponding objective value.

Table 2-3 Solver Time (sec.) v.s. Number of Wind Farms

wind farms No.	5	10	15	20	25	30
case118	3.1	4.5	3.5	2.8	3.6	2.6
case1888	36.9	41.8	39.5	45.0	46.4	49.6

is introduced in the affine reserve procurement process. Table 2-4 summarizes the reserve allocation by the proposed method on the 118-bus system during 4:00∼5:00 a.m.. Five units are delegated as spinning reserve units with total reserve 270.44 MW. These units are of two different reserve procurement prices. When wind power forecasting error occurs, to minimize the reserve procurement costs, the cheaper units should be completely utilized before resorting to any more expensive ones. Hence the reserve procurement costs corresponding to the optimal recourse policy is represented by the red dotted line in Fig. 2-11. However, the affine policy

in the proposed model initiates all the reserve units simultaneously with different participation factors. The reserve procurement costs corresponding to the affine policy is shown with the blue line in Fig. 2-11. Clearly, there is a gap between the two lines due to the discrepancy of reserve prices. When beta distribution is used to simulate the forecasting error, the expected costs in 24h are summarized in Table 2-5. The affine policy has increased the reserve procurement costs by \$873.35 which accounts for 4.88% of the reserve procurement costs and 0.08% of the total operation costs. The above numerical study confirms the observation made in [74] and [76] that the simple affine policy (2-10) introduces minor conservatism.

Table 2-4 Summary of the Reserve Allocation of 118-bus System during 4:00∼5:00 a.m.

unit no.	participation factor	upward reserve (MW)	procurement price (\$/MW)	total reserve
4	0.319	86.153	14.176	270.441 MW
27	0.048	13.105	9.173	
28	0.342	92.528	9.173	
36	0.078	21.220	14.176	
40	0.212	57.435	14.176	

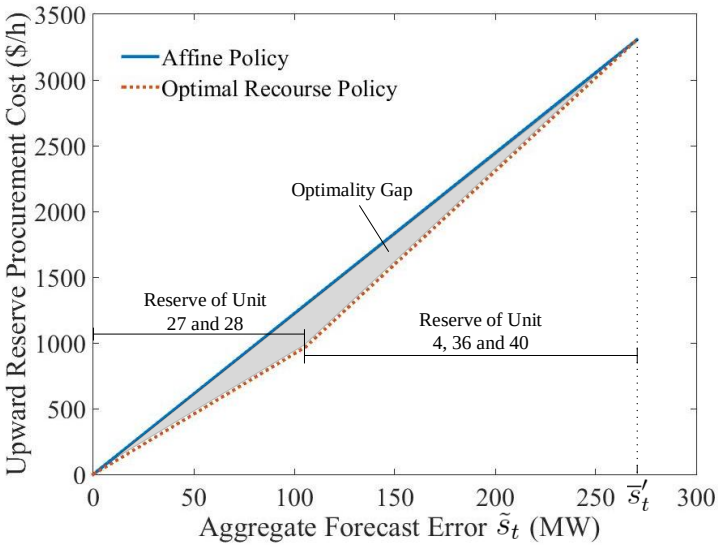


Figure 2-11 Illustration of optimality gap between affine reserve procurement policy and optimal recourse reserve procurement policy on 118-bus system at 5 a.m.

2.4.4 Comparing Different Methods to Deal with Uncertainties

To compare the proposed DRO with the other two methods to deal with uncertainties, we also implement the RO and SP in our problem formulation in application mode-I. The simple affine policy (2-10) is used in three methods to provide a clear baseline for the comparison.

Table 2-5 Comparion between Affine Reserve Procurement and fully Adaptive Reserve Procurement

Reserve Procurement Cost by Affine Policy	18756.14	Total operation Cost	1128619.02
Reserve Procurement Cost by Optimal Recourse Policy	17882.79	Percentage Gap w.r.t. Reserve Procurement Cost	4.88%
Gap	873.35	Percentage Gap w.r.t. Total Cost	0.08%

The uncertainty set employed in RO is the estimated support of the random variable. The SP assumes the random variables follow the normal distribution with mean and variance being the sample mean and sample variance of the historical data. Laplace distribution is taken as the underlying true distribution for the random variables. After solving each problem, MCS is employed to test the practical performance, including operation costs and reliability guarantee, of the corresponding strategies. Fig. 2-12 compares the operation costs of different methods, including RO, SP and the proposed DRO with $10^3, 10^4$ and 10^5 historical data points. The operation costs of the RO are the highest whereas those of the SP are the lowest. The DROs with different amount of data are intermediates between RO and SP. The costs of DRO go close to those of SP as more historical data is available. The probability of load shedding and wind curtailment for different methods are further compared in Fig. 2-13. RO has the far-more-than-required level of reliability due to the ignorance of probabilistic information. The SP, however, does not guarantee the required reliability level because the underlying true distribution differs from the normality assumption made in the method. The proposed DRO approach always provide safe reliability guarantee and the practical reliability level approaches the pre-specified reliability indices as more data is available. We summarize the features of the three methods in Table 2-6.

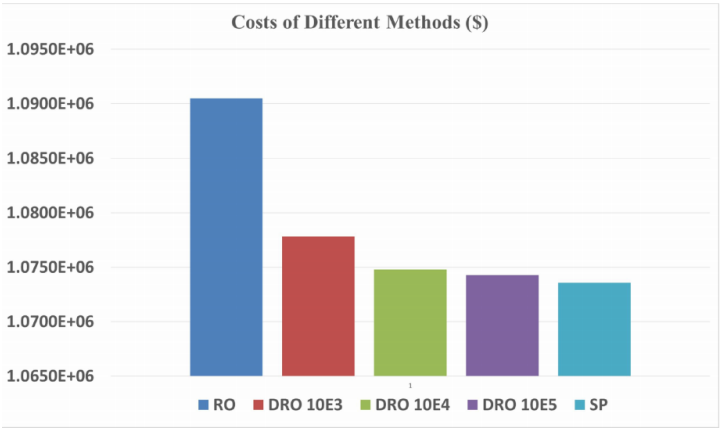
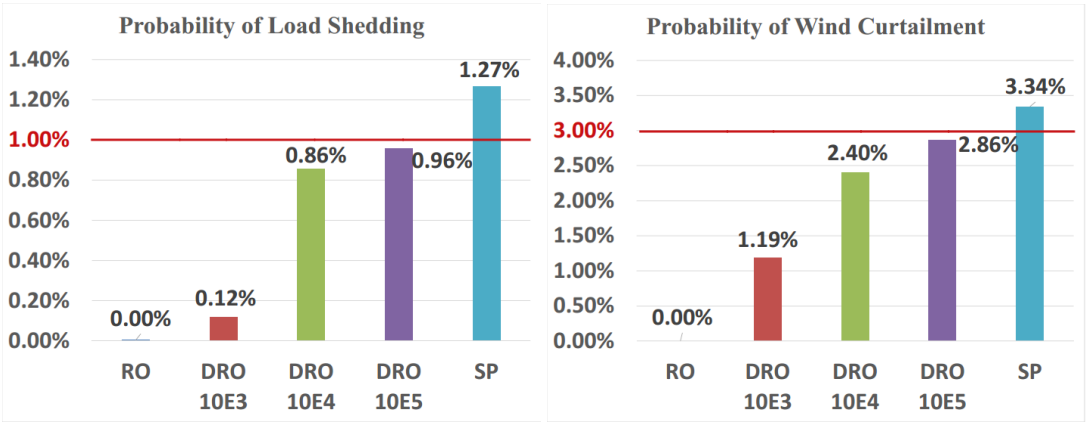


Figure 2-12 Comparison of operation costs by MCS on 118-bus system using different methods, including RO, SP and the proposed DRO with $10^3, 10^4$ and 10^5 historical data points.



(a) Load Shedding (b) Wind Curtailment

Figure 2-13 Comparison of probability of load shedding and wind curtailment by MCS on 118-bus system using different methods, including RO, SP and the proposed DRO with $10^3, 10^4$ and 10^5 historical data points.

Table 2-6 Comparison among Different Methods to Deal with Uncertainties

	the proposed DRO	RO	SP
exploit probabilistic information	yes	no	yes
presumption on distribution	no	no	yes
reliability guarantee when true distribution is unknown	yes	yes	no
data-exploiting: the more data, the less conservative	yes	no	no

2.5 Conclusion and Discussion

This chapter applies the idea of distributionally robust optimization to the UC under uncertainty. Based on the non-parametric inference theory, an ambiguity set that contains the true probability distribution of uncertainties is constructed from observed historical data. The proposed model considers the worst-case distribution in the ambiguity set thereby achieves operational robustness. Moreover, the proposed ambiguity set shrinks to the true distribution as the amount of historical data increases. Therefore, the conservatism of the solution can be reduced by incorporating more data. In addition, the scale of the optimization problem remains unchanged when using more data.

Recently, dynamic uncertainty set [77, 86] and data-driven uncertainty set [87] have been proposed to reduce the conservatism of RO. Compared with the dynamic uncertainty set method which explicitly models the correlation and dynamics of uncertainties, the proposed approach to deal with uncertainty is still static in nature. But the solution of the proposed method is much more simple and direct. The treatment of the distributionally robust chance constraints in this chapter is basically the same as the data-driven uncertainty set with probabilistic guarantee [87], but different hypothesis tests are employed. More importantly, the essential difference between the ambiguity-set-based DRO and the uncertainty-set-based RO is that the former minimizes expected costs w.r.t the worst-case distribution whereas the latter minimizes the worst-case costs. Therefore, DRO exploits much more probabilistic information than RO. This chapter and reference [36] share some similarities as both works employ the idea of DRO and affine policy. The ambiguity sets used in both papers allow reformulation of DRO problems into tractable MILPs. Unlike that in [36], the ambiguity set proposed in this chapter has sound statistical foundation and data-exploiting ability. Meanwhile, the proposed approach is more scalable than that in [36] since the scale of MILP in this chapter remains unchanged as the amount of historical data and the number of uncertain renewable sources increase.

However, the proposed method also has limitations. The ambiguity set for $\mathbb{P}_t^l$ only incoporates the information from its marginal distributions, which brings additional conservatism to the chance constraints for line flow. One future direction to remedy this issue is to consider a Wasserstein-metric-based ambiguity set which can effectively deal with multivariate distributions [88]. In addition, the simple affine policy is merely a conservative approximation to the optimal recourse action. Therefore, another direction is to apply the proposed ambiguity set and reformulation technique to evaluate the operation risk in the risk-constrained UC [89]. A fully adaptive data-driven UC method can then be obtained by leveraging the results from this chapter and reference [89].

3 Data-driven Distributionally Robust Energy-Reserve-Storage Dispatch

3.1 Introduction

Large-scale integration of renewable energy has brought a high level of variability and uncertainty into power system operation, which poses a great challenge to system schedule and dispatch. Energy storage systems (ESS) are recognized as underpinning technologies to meet such challenge due to their ability to provide time-varying energy management and alleviate the intermittence of renewable generation [90]. To minimize operation costs while guarantee system reliability under variability and uncertainty, the operation of ESSs must be integrated into the conventional economic dispatch and reserve scheduling problems.

The investigation of the optimal operation of power system with ESSs requires the multi-period optimal power flow (OPF) models because the operation of ESS is strongly coupled over time by charge/discharge dynamics. Both DC and AC power flow models are applied to this problem. Jabr *et al* developed a robust multi-period DC OPF with ESS to address the uncertainties of renewable generation [75]. DC flow based multi-period OPF is also employed in [91] to optimize storage allocation and portfolio. Since DC power flow neglects voltage magnitude and reactive power, its results might be unreasonable for practical operation. Thus, full AC formulations of multi-period OPF are also introduced to address the optimal operation of distribution networks in [92] and [93]. But due to the non-convexity of the problem formulation, only local optimality is guaranteed. By extending the seminar work of Lavaei and Low [65], semi-definite program (SDP) relaxations of AC flow based multi-period OPF are discussed in [94] where global optimal solutions are achievable in some cases.

Conventionally, spinning reserve is not explicitly handled in economic dispatch but treated separately in the reserve scheduling problem [63, 95]. Due to the ever-increasing level of uncertainty and the commercialization of spinning reserve as an auxiliary service, co-optimization of energy and reserve becomes a trend in recent literature [64, 96, 97]. When the load and renewable generation deviate from the predicted values, spinning reserve needs to be utilized to maintain real-time power balance. The process of reserve procurement is usually handled in two different approaches. The first is the affinely adjustable approach [75, 98] where a generator uses its reserve according to the associated participation factor. The second is the fully adjustable approach [86, 97] in which the reserve procurement is treated as a sub-level optimization problem after the realization of uncertainties. The advantages of affinely adjustable approach are its compatibility with existing automatic generation systems (AGC) and the numerical tractability of the optimization model. Nevertheless, the affine policy is more

restrictive thus brings some conservatism compared with full recourse strategy.

Besides above-mentioned aspects related to problem formulation, a more prominent issue is how to deal with uncertainties. Stochastic programming (SP) [21, 22], robust optimization (RO) [74, 75, 86, 96–99] and distributionally robust optimization (DRO) [38, 42, 63, 64] have been employed to tackle uncertainties in power system operation. SP assumes operational uncertainties follow a pre-specified probability distribution and characterizes the uncertainties by scenarios sampled from that distribution. In contrast to SP, RO does not require any probabilistic information of the uncertainties. Instead, randomness is represented by a deterministic uncertainty set, and RO seeks strategies that are immune against all realizations of the uncertainty set. In practice, the probability distribution of uncertainties truly exists but must be estimated from historical data and is therefore itself uncertain. To bridge the gap between the specificity of SP and conservatism of RO, DRO assumes that the true distribution lies in an ambiguity set and immunizes the operation strategies against all distributions in the ambiguity set. The ambiguity set employed in [38, 42, 63, 64] is the set of all probability distributions sharing given mean and covariance. Paper [38] further assumes the unimodality of the distribution to reduce conservatism. However, the DRO approach discussed in [38, 63, 64] has the following drawbacks. First, the ambiguity set characterized only by the first two moments is in fact very large thus the method is still very conservative. Second, moments also need to be estimated from historical data therefore are uncertain as well. Third, the problems are finally cast into a sequence of semi-definite programmings which are very computationally intensive.

In this chapter, we propose novel formulation and method for co-optimization of energy, reserve, and storage under the spirit of DRO [78]. The contributions are threefold:

1. **Problem formulation:** We extend the robust multi-period OPF formulation [75] to co-dispatch of energy, reserve, and storage. Distributionally robust chance constraints (DRCC) are employed to provide explicit reliability guarantee for reserve and transmission adequacy. The objective is to minimize the expected operation costs w.r.t. the worst-case distribution in the constructed ambiguity set, which provides robustness for economical system operation.
2. **Data-driven and data-exploiting features:** The proposed method is data-driven in the sense that the ambiguity set for DRO is constructed from historical data without any prior knowledge about the distribution. The method can automatically extract and exploit the probabilistic information contained in the data set. The more historical data is available, the less conservative the solution is.
3. **Efficient solution approach:** The problem is finally formulated as a mixed integer linear programming (MILP) for which off-the-shelf solvers are available. The scale of the MILP remains unchanged as the number of available data increases. Computational issues, including the elimination of inactive line capacity constraints and convex

relaxation of binary variables, are considered in the solution approach to significantly improve the numerical tractability.

Nomenclature

$\mathcal{B}, \mathcal{L}, \mathcal{T}$	Set of all buses, lines and time periods.
c_i^g	Generation price of the i_{th} generator.
c_i^{sc}/c_i^{sd}	Charge/discharge price of the i_{th} storage.
$d_i^{g,up}/d_i^{g,dn}$	Upward/downward reserve availability price of the i_{th} generator.
$d_i^{s,up}/d_i^{s,dn}$	Upward/downward reserve availability price of the i_{th} storage.
$f_i^{g,up}/f_i^{g,dn}$	Upward/downward reserve utilization price of the i_{th} generator.
$f_i^{s,up}/f_i^{s,dn}$	Upward/downward reserve utilization price of the i_{th} storage.
$\underline{P}_i^g/\overline{P}_i^g$	Lower/upper limits of output power of the i_{th} generator.
$\overline{R}_i^{g,up}/\overline{R}_i^{g,dn}$	Ramp up/down rate limits of the i_{th} generator.
$\overline{P}_i^{sc}/\overline{P}_i^{sd}$	Upper limits for the charge/discharge power of the i_{th} storage.
C_l	Capacity of transmission line l.
π_{gi}^l/π_{si}^l	Load shift factor from generator/storage/load i to line l.
π_{di}^l	Load shift factor from load i to line l.
η_i^{sc}/η_i^{sd}	Charge/discharge efficiency of the i_{th} storage.
$\tilde{\epsilon}_{it}$	Random forecasting error of composite load at the i_{th} bus in time t.
$\hat{p}_{it}^d$	Forecasted composite load at the i_{th} bus in time t.
$\tilde{p}_{it}^d$	Random composite load at the i_{th} bus in time t.
p_{it}^g	Output power setting point of the i_{th} generator in time t.
$\tilde{p}_{it}^g$	Random actual output power of the i_{th} generator in time t.
$\tilde{s}_{lt}$	Random line power flow on line l in time t.
$\tilde{p}_{it}^s$	Random actual output power of the i_{th} storage in time t.
e_{i0}^s	Initial stored energy of the i_{th} storage.
α_{it}^g	Participation factor of the i_{th} generator in time t.
α_{it}^s	Participation factor of the i_{th} storage in time t.
p_{it}^{sc}/p_{it}^{sd}	Charge/discharge power setting point of the i_{th} storage in time t.
$r_{it}^{g,up}/r_{it}^{g,dn}$	Upward/downward reserve of the i_{th} generator in time t.
$r_{it}^{s,up}/r_{it}^{s,dn}$	Upward/downward reserve of the i_{th} storage in time t.
$\mathbb{P}$	A probability measure/distribution.
$\mathbb{E}_{\mathbb{P}}$	Expectation respect to probability measure $\mathbb{P}$.
$\mathcal{P}_0(\Xi)$	Set of all probability measures with support Ξ.
$I_{\Xi}(\cdot)$	Indicator function of set Ξ, i.e. $I_{\Xi}(x) = 1$ when $x \in \Xi$ and 0 otherwise.
$(x)^+$	$\max\{x, 0\}$.

3.2 Distributionally Robust Optimization

3.2.1 Basic Concepts

In power system, some operational strategy x is needed to minimize the cost function $f(x,\tilde{\xi})$ while satisfy the technical and security constraints $g(x,\tilde{\xi}) \leq 0$ where both the cost function and constraints are affected by some uncertainty represented by random variable $\tilde{\xi}$. In practice, the probability distribution of $\tilde{\xi}$ is unknown and only some historical data is available. Theoretically, the precise description of the probability distribution cannot be obtained from finite sample data. Therefore, the probability distribution itself is uncertain. However, the historical data does provide us some reliable information about the distribution, and based on these information we can construct an ambiguity set $\mathcal{P}$, i.e. a set of probability distributions consistent with the observed historical data. Hence the *distributionally robust optimization* seeks decisions that are immune against all distributions and perform best in view of the worst-case distribution from the ambiguity set, i.e.

$$\begin{aligned} &\min_{x} \max_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\{f(x,\tilde{\xi})\} \\ &\text{s.t. } \mathbb{P}\{g(x,\tilde{\xi}) \leq 0\} \geq 1-\beta, \forall \mathbb{P} \in \mathcal{P}. \end{aligned} \tag{3-1}$$

The performance and numerical tractability of the above problem largely rely on the structure of the ambiguity set. A desirable ambiguity set should possess the following properties: 1) $\mathcal{P}$ contains the underlying true probability distribution; 2) to reduce conservatism, $\mathcal{P}$ can be made as small as possible by incorporating more observed data; 3) the structure of $\mathcal{P}$ allows the reformulation of distributionally robust optimization problem (3-1) into tractable deterministic problem.

3.2.2 Construction of Ambiguity Set

In this Chapter, we employ the same approach as in Chapter 2 to construct the ambiguity set $\mathcal{P}$ based on confidence bands for cumulative distribution function (CDF). Consider a one-dimensional random variable $\tilde{\xi}$ whose probability distribution is unknown whereas the ascendingly ordered sample set $S = \{\hat{\xi}^{(1)}, \hat{\xi}^{(2)}, \cdots, \hat{\xi}^{(n)}\}$ is available. Let $F(x) = \mathbb{P}^*\{\xi \leq x\}$ be the CDF of true distribution $\mathbb{P}^*$. The $1-\alpha$ confidence bands for $F(x)$ is a pair of sample-dependent functions $\underline{P}(x)$ and $\overline{P}(x)$ for which $\underline{P}(x) \leq F(x) \leq \overline{P}(x)$, $\forall x \in \mathbb{R}$ with probability $1-\alpha$ over the choice of sample set S. One approach to obtain the confidence bands for $F(x)$ is the Dirichlet method [71] summarized blow:

Lemma 3.1： Let $S = \{\hat{\xi}^{(1)}, \hat{\xi}^{(2)}, \cdots, \hat{\xi}^{(n)}\}$ be the ascendingly ordered sample set of random variable ξ generated independently according to true distribution $\mathbb{P}^*$ with continuous CDF $F(x)$. $B^{\alpha}_{k,n}$ denote the α-quantile of the $\beta(k, n+1-k)$ distribution. For given n and α,

define $\underline{p}_k = B_{k,n}^{\tilde{\alpha}/2}$ and $\overline{p}_k = B_{k,n}^{1-\tilde{\alpha}/2}$ where

$$\tilde{\alpha} = \exp\left(-c_1(\alpha) - c_2(\alpha)\sqrt{\ln[\ln(n)]} - c_3(\alpha)[\ln(n)]^{c_4(\alpha)}\right) \tag{3-2}$$

with $c_1(\alpha) = -2.75 - 1.04\ln(\alpha)$, $c_2(\alpha) = 4.76 - 1.20\alpha$, $c_3(\alpha) = 1.15 - 2.39\alpha$, and $c_4(\alpha) = -3.96 + 1.72\alpha^{0.171}$. Add $\hat{\xi}^{(0)} = -\infty$ and $\hat{\xi}^{(n+1)} = \infty$ to the ascending sequence of the sample set S, and define $\underline{p}_0 = 0$ and $\overline{p}_{n+1} = 1$. Then

$$\underline{P}(x) = \max\{\underline{p}_k : \hat{\xi}^{(k)} \leq x\} \tag{3-3}$$

$$\overline{P}(x) = \min\{\overline{p}_k : \hat{\xi}^{(k)} \leq x\} \tag{3-4}$$

are the $1-\alpha$ confidence bands for $F(x)$. □

Note that $\underline{P}(x)$ and $\overline{P}(x)$ have the following properties: 1) they are stair-step functions that take values $\underline{p}_k$ and $\overline{p}_k$ at $\hat{\xi}^{(k)}$, respectively; 2) the empirical CDF $\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{\hat{\xi}^{(i)} \leq x\}}$ is lower and upper bounded by $\underline{P}(x)$ and $\overline{P}(x)$, i.e. $\underline{P}(x) \leq \hat{F}(x) \leq \overline{P}(x)$; 3) as the size of the sample set $n \to \infty$, $\sup|\underline{P}(x) - \overline{P}(x)| \to 0$. In other words, $\underline{P}(x)$ and $\overline{P}(x)$ represent the reliable information that can be extracted from finite samples and the information becomes more and more accurate as the size of the sample set grows.

Based on the confidence bands for CDF, the ambiguity set $\mathcal{P}$ employed in this chapter takes the form

$$\mathcal{P} = \left\{\mathbb{P} \in \mathcal{P}_0([\underline{\xi}, \overline{\xi}]) \middle| \mathbb{P}\{\tilde{\xi} \leq \hat{\xi}^{(k)}\} \in [\underline{p}_k, \overline{p}_k],\ k = 1, \cdots, n\right\} \tag{3-5}$$

where $\mathcal{P}_0([\underline{\xi}, \overline{\xi}])$ denotes the set of all probability measures whose supports are the interval $[\underline{\xi}, \overline{\xi}]$. The proposed **CDF-based ambiguity set** $\mathcal{P}$ is designed to encode the information from confidence bands for CDF and does not assume any prior knowledge about the distribution type. Due to the convergence property of the confidence bands, the ambiguity set $\mathcal{P}$ is made smaller and smaller by incorporating more and more historical data. Moreover, the structure defined in (3-5) allows very efficient reformulation of distrituionally robust optimization problems, which will be analyzed in section 3.4.

3.3 Problem Formulation

The problem formulation of distributionally robust energy-reserve-storage dispatch proposed in this chapter is directly extended from Jabr's robust multi-period OPF with storage and renewables [75]. The extension is made from two aspects. First, spinning reserve is explicitly handled in the formulation, and the availability and utilization costs of the spinning reserve are thus reflected in the objective function. Second, instead of robust optimization, the problem is formulated as a distributionally robust optimization problem as (3-1).

The uncertainties of system operation mainly originate from the the load and renewable forecasting errors uniformly represented by the forecasting errors of composite load:

$$\tilde{p}_{it}^{d} = \hat{p}_{it}^{d} + \tilde{\epsilon}_{it}. \tag{3-6}$$

To maintain real-time power balance under uncertainties, the proposed formulation inherits the affinely adjustable approach from [75] where both the conventional generators and the energy storages participate in the frequency regulation according to the associated participation factors:

$$\tilde{p}_{it}^{g} = p_{it}^{g} + \alpha_{it}^{g} \sum_{k\in\mathcal{B}} \tilde{\epsilon}_{kt}, \tag{3-7}$$

$$\tilde{p}_{it}^{s} = p_{it}^{sd} - p_{it}^{sc} + \alpha_{it}^{s} \sum_{k\in\mathcal{B}} \tilde{\epsilon}_{kt}. \tag{3-8}$$

Although the affinely adjustable approach [74, 75, 98, 99] is only conservative approximation to the fully adjustable approach [64, 86], affine policy is directly compatible to the AGC and numerically more tractable. To reduce the dimension of the random variables, we further define

$$\tilde{\phi}_t = \sum_{k\in\mathcal{B}} \tilde{\epsilon}_{kt} \tag{3-9}$$

$$\tilde{\theta}_{lt} = \sum_{k\in\mathcal{B}} \pi_{dk}^{l} \tilde{\epsilon}_{kt}. \tag{3-10}$$

Under above definitions, we have $\tilde{p}_{it}^{g} = p_{it}^{g} + \alpha_{it}^{g}\tilde{\phi}_t$ and $\tilde{p}_{it}^{s} = p_{it}^{sd} - p_{it}^{sc} + \alpha_{it}^{s}\tilde{\phi}_t$. Furthermore, by using equation (3-6)∼(3-10), the power flow on line l at time t takes the form as in equation (3-11).

$$\tilde{s}_{lt} = \sum_{i\in\mathcal{G}} \pi_{gi}^{l}\tilde{p}_{it}^{g} + \sum_{i\in\mathcal{S}} \pi_{si}^{l}\tilde{p}_{it}^{s} - \sum_{i\in\mathcal{B}} \pi_{di}^{l}\tilde{p}_{it}^{d} \tag{3-11a}$$

$$= \sum_{i\in\mathcal{G}} \pi_{gi}^{l}(p_{it}^{g} + \alpha_{it}^{g}\tilde{\phi}_t) + \sum_{i\in\mathcal{S}} \pi_{si}^{l}(p_{it}^{sd} - p_{it}^{sc} + \alpha_{it}^{s}\tilde{\phi}_t) - \sum_{i\in\mathcal{B}} \pi_{di}^{l}\hat{p}_{it}^{d} - \tilde{\theta}_{lt} \tag{3-11b}$$

Let $\mathbb{P}_t^l$ denote the probability distribution for 2-dimensional random variable $(\tilde{\phi}_t, \tilde{\theta}_{lt})$ with marginal distributions $\mathbb{P}_t^{\phi}$ and $\mathbb{P}_t^{\theta^l}$. Using the historical data of $\tilde{\phi}_t$ and $\tilde{\theta}_{lt}$, we can construct the ambiguity sets as defined in (3-5) for $\mathbb{P}_t^{\phi}$ and $\mathbb{P}_t^{\theta_l}$, denoted as $\mathcal{P}_t^{\phi}$ and $\mathcal{P}_t^{\theta_l}$, respectively. Then the ambiguity set of joint distribution $\mathbb{P}_t^l$ is defined as $\mathcal{P}_t^l = \left\{\mathbb{P}_t^l \in \mathcal{P}_0(\mathcal{R}^2) \middle| \mathbb{P}_t^{\phi} \in \mathcal{P}_t^{\phi}, \mathbb{P}_t^{\theta_l} \in \mathcal{P}_t^{\theta_l}\right\}$.

Under the affinely adjustable framework described above, the real-time power balance is guaranteed by

$$\sum_{i\in\mathcal{G}} p_{it}^{g} + \sum_{i\in\mathcal{S}} (p_{it}^{sd} - p_{it}^{sc}) = \sum_{i\in\mathcal{B}} \hat{p}_{it}^{d} \tag{3-12}$$

$$\sum_{i\in\mathcal{G}} \alpha_{it}^{g} + \sum_{i\in\mathcal{S}} \alpha_{it}^{s} = 1 \tag{3-13}$$

$$\alpha_{it}^{g} \geq 0,\ \alpha_{it}^{s} \geq 0, \tag{3-14}$$

where the setting values of generator power, storage charge/discharge power and generator/storage participator factors need to be dynamically adjusted to minimize costs while ensure system security. The followed constraints (3-15)∼(3-17)

$$\underline{P}_{i}^{g} + r_{it}^{g,dn} \leq p_{it}^{g} \leq \overline{P}_{i}^{g} - r_{it}^{g,up} \tag{3-15}$$

$$r_{it}^{g,dn} \geq 0,\ r_{it}^{g,up} \geq 0 \tag{3-16}$$

$$-\overline{R}_{i}^{g,dn} + r_{i(t+1)}^{g,dn} + r_{it}^{g,up} \leq p_{gi}(t+1) - p_{it}^{g} \leq \overline{R}_{i}^{g,up} - r_{i(t+1)}^{g,up} - r_{it}^{g,dn}. \tag{3-17}$$

together ensure the procurability of upward and downward spinning reserve from conventional generators considering generator capacity limits (3-15) and ramp rate limits (3-17). By introducing binary decision variable ω_{it}, the following constraints

$$0 \leq p_{it}^{sc} \leq \overline{P}_{i}^{sc}\omega_{it} \tag{3-18a}$$

$$0 \leq p_{it}^{sd} \leq \overline{P}_{i}^{sd}(1-\omega_{it}) \tag{3-18b}$$

$$\omega_{it} \in \{0,1\},\ i \in \mathcal{S},\ t \in \mathcal{T} \tag{3-18c}$$

not only set the storage charge/discharge power limits but also avoid simultaneous charging and discharging. The followed constraints (3-19)∼(3-22)

$$-\overline{P}_{i}^{sc} + r_{it}^{s,dn} \leq p_{it}^{sd} - p_{it}^{sc} \leq \overline{P}_{i}^{sd} - r_{it}^{s,up} \tag{3-19}$$

$$r_{it}^{s,dn} \geq 0,\ r_{it}^{s,up} \geq 0 \tag{3-20}$$

$$e_{i0}^{s} + \sum_{\tau=1}^{t} \left(\eta_{i}^{sc} p_{i\tau}^{sc} - \frac{1}{\eta_{i}^{sd}} p_{i\tau}^{sd} + \eta_{i}^{sc} r_{it}^{s,dn} \right) \Delta t \leq \overline{E}_{si} \tag{3-21}$$

$$e_{i0}^{s}+\sum_{\tau=1}^{t}\left(\eta_{i}^{sc}p_{i\tau}^{sc}-\frac{1}{\eta_{i}^{sd}}p_{i\tau}^{sd}-\frac{1}{\eta_{i}^{sd}}r_{it}^{s,up}\right)\Delta t\geq\underline{E}_{si} \tag{3-22}$$

together guarantee the procurability of spinning reserve from energy storages by considering the charge/discharge power limits (3-19) and the stored energy upper/lower limits (3-21)(3-22). The following cycling constraint

$$e_{i0}^{s}+\sum_{\tau=1}^{T}\left(\eta_{i}^{sc}p_{i\tau}^{sc}-\frac{1}{\eta_{i}^{sd}}p_{i\tau}^{sd}\right)\Delta t=e_{i0}^{s} \tag{3-23}$$

sets the final stored energy to be the initial values.

In addition, the adequacy of downward/upward spinning reserve and line capacity is ensured by distributionally robust chance constraint (DRCC)

$$\mathbb{P}_{t}^{\phi}\left\{\begin{array}{l}-r_{it}^{g,dn}\leq\alpha_{it}^{g}\tilde{\phi}_{t},\ \forall i\in\mathcal{G}\\-r_{it}^{s,dn}\leq\alpha_{it}^{s}\tilde{\phi}_{t},\ \forall i\in\mathcal{S}\end{array}\right\}\geq1-\beta_{1},\forall\mathbb{P}_{t}^{\phi}\in\mathcal{P}_{t}^{\phi} \tag{3-24}$$

$$\mathbb{P}_{t}^{\phi}\left\{\begin{array}{l}\alpha_{it}^{g}\tilde{\phi}_{t}\leq r_{it}^{g,up},\ \forall i\in\mathcal{G}\\\alpha_{it}^{s}\tilde{\phi}_{t}\leq r_{it}^{s,up},\ \forall i\in\mathcal{S}\end{array}\right\}\geq1-\beta_{2},\forall\mathbb{P}_{t}^{\phi}\in\mathcal{P}_{t}^{\phi} \tag{3-25}$$

and

$$\mathbb{P}_{t}^{l}\left\{\begin{array}{l}-C_{l}\leq\sum\limits_{i\in\mathcal{G}}\pi_{gi}^{l}(p_{it}^{g}+\alpha_{it}^{g}\tilde{\phi}_{t})\\+\sum\limits_{i\in\mathcal{S}}\pi_{si}^{l}(p_{it}^{sd}-p_{it}^{sc}+\alpha_{it}^{s}\tilde{\phi}_{t})\\\qquad-\sum\limits_{i\in\mathcal{B}}\pi_{di}^{l}\hat{p}_{it}^{d}-\tilde{\theta}_{lt}\leq C_{l}\end{array}\right\}\geq1-\gamma,\forall\mathbb{P}_{t}^{l}\in\mathcal{P}_{t}^{l} \tag{3-26}$$

where the parameter β_1, β_2 and γ are pre-specified allowable probability for renewable curtailment, load shedding and transmission line overload, respectively. In order for (3-24)∼(3-24) to be feasible, it requires $\gamma\geq\beta_1+\beta_2$. We will shed more light on the DRCC (3-24)∼(3-26) by looking at their deterministic counterparts in section 3.4.1.

The objective function consists of the costs of conventional generation (3-27a), storage charge/discharge (3-27b), upward/downward reserve availability of generators (3-27c), upward/downward reserve availability of storages (3-27d) and the expected costs of reserve utilization (3-27e), formally stated as

$$\begin{aligned}&F(\boldsymbol{p}^{g},\boldsymbol{p}^{sd},\boldsymbol{p}^{sc},\boldsymbol{r}^{g,up},\boldsymbol{r}^{g,dn},\boldsymbol{r}^{s,up},\boldsymbol{r}^{s,dn},\boldsymbol{\alpha}^{g},\boldsymbol{\alpha}^{s})\\=&\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{G}}c_{i}^{g}p_{it}^{g} &\text{(3-27a)}\\&+\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{S}}(c_{i}^{sd}p_{it}^{sd}+c_{i}^{sc}p_{it}^{sc}) &\text{(3-27b)}\\&+\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{G}}(d_{i}^{g,up}r_{it}^{g,up}+d_{i}^{g,dn}r_{it}^{g,dn}) &\text{(3-27c)}\end{aligned}$$

$$+\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{S}}(d_i^{s,up}r_{it}^{s,up}+d_i^{s,dn}r_{it}^{s,dn}) \tag{3-27d}$$

$$+\sum_{t\in\mathcal{T}}\max_{\mathbb{P}_t^\phi\in\mathcal{P}_t^\phi}\mathbb{E}_{\mathbb{P}_t^\phi}\{Q_t(\boldsymbol{\alpha}_t^g,\boldsymbol{\alpha}_t^s,\tilde{\phi}_t)\}. \tag{3-27e}$$

In light of (3-1), the expectation in (3-27e) is evaluated w.r.t. the worst-case distribution in the ambiguity set, which guarantees that the obtained strategy can perform well in the absence of precise knowledge about the underlying true probability distribution. The explicit formula for reserve utilization costs $Q_t(\boldsymbol{\alpha}_g(t),\boldsymbol{\alpha}_s(t),\tilde{\phi}_t)$ is given by

$$\begin{aligned}&Q_t(\boldsymbol{\alpha}_t^g,\boldsymbol{\alpha}_t^s,\tilde{\phi}_t)=\\&\sum_{i\in\mathcal{G}}\left(f_i^{g,up}(\alpha_{it}^g\min\{\tilde{\phi}_t,\overline{\phi}_t'\})^+ + f_i^{g,dn}(-\alpha_{it}^g\max\{\tilde{\phi}_t,\underline{\phi}_t'\})^+\right)\\&+\sum_{i\in\mathcal{S}}\left(f_i^{s,up}(\alpha_{it}^s\min\{\tilde{\phi}_t,\overline{\phi}_t'\})^+ + f_i^{s,dn}(-\alpha_{it}^s\max\{\tilde{\phi}_t,\underline{\phi}_t'\})^+\right).\end{aligned} \tag{3-28}$$

where $[\underline{\phi}',\overline{\phi}']$, that will be analyzed in section 3.4.1, is the dispatchable range of total forecasting error determined by the DRCC (3-24)$\sim$(3-26).

To sum up, the proposed distributionally robust energy-reserve-storage dispatch (ERSD) problem takes the form:

$$\begin{aligned}&\min F(\boldsymbol{p}^g,\boldsymbol{p}^{sd},\boldsymbol{p}^{sc},\boldsymbol{r}^{g,up},\boldsymbol{r}^{g,dn},\boldsymbol{r}^{s,up},\boldsymbol{r}^{s,dn},\boldsymbol{\alpha}^g,\boldsymbol{\alpha}^s)\\&\text{s.t. }\ (3-12)\sim(3-26).\end{aligned} \tag{3-29}$$

3.4 Solution Approach

The major obstacles to solve problem (3-29) are the DRCCs (3-24)$\sim$(3-26) and the worst-case expectation in the objective function (3-27). In this section, we show the DRCCs can be replaced by some deterministic linear constraints and the worst-case expectation can be evaluated by a linear programming. Therefore, the problem (3-29) can be cast into a MILP.

3.4.1 Reformulation of Distributionally Robust Chance Constraints

Let $\underline{\Phi}_t(x)$ and $\overline{\Phi}_t(x)$ be the confidence bands for the CDF of random variable $\tilde{\phi}_t$, and further define $\underline{\phi}_t'=\overline{\Phi}_t^{-1}(\beta_1)$ and $\overline{\phi}_t'=\underline{\Phi}_t^{-1}(1-\beta_2)$. Then DRCC (3-24) is satisfied if

$$\begin{cases}-r_{it}^{g,dn}\le\alpha_{it}^g\underline{\phi}_t',\ \forall i\in\mathcal{G}\\-r_{it}^{s,dn}\le\alpha_{it}^s\underline{\phi}_t',\ \forall i\in\mathcal{S},\end{cases} \tag{3-30}$$

which can be easily seen from $\mathbb{P}_t^\phi\{\tilde{\phi}_t<\underline{\phi}_t'\}\le\overline{\Phi}_t(\underline{\phi}_t')=\beta_1,\ \forall\mathbb{P}_t^\phi\in\mathcal{P}_t^\phi$. Similarly, DRCC (3-25) is satisfied if

$$\begin{cases}\alpha_{it}^g\overline{\phi}_t'\le r_{it}^{g,up},\ \forall i\in\mathcal{G}\\\alpha_{it}^s\overline{\phi}_t'\le r_{it}^{s,up},\ \forall i\in\mathcal{S},\end{cases} \tag{3-31}$$

since $\mathbb{P}_t^{\phi}[\tilde{\phi}_t > \overline{\phi}_t'] \leq 1-\underline{\Phi}_t(\overline{\phi}_t') = \beta_2$. In addition, let $\underline{\Theta}_t^l(x)$ and $\overline{\Theta}_t^l(x)$ be the confidence bands for the CDF of random variable $\tilde{\theta}_{lt}$, and $\underline{\theta}_{lt}' = (\overline{\Theta}_t^l)^{-1}(\frac{\gamma-\beta_1-\beta_2}{2})$ and $\overline{\theta}_{lt}' = (\underline{\Theta}_t^l)^{-1}(1-\frac{\gamma-\beta_1-\beta_2}{2})$. Then we have $\forall \mathbb{P}_t^l \in \mathcal{P}_t^l$,

$$\begin{aligned}
&\mathbb{P}_t^l\left\{\tilde{\phi}_t \notin [\underline{\phi}_t', \overline{\phi}_t'] \text{ or } \tilde{\theta}_{lt} \notin [\underline{\theta}_{lt}', \overline{\theta}_{lt}']\right\} \\
\leq &\mathbb{P}_t\{\tilde{\phi}_t < \underline{\phi}_t'\} + \mathbb{P}_t\{\tilde{\phi}_t > \overline{\phi}_t'\} + \mathbb{P}_t\{\tilde{\theta}_{lt} < \underline{\theta}_{lt}'\} + \mathbb{P}_t\{\tilde{\theta}_{lt} > \overline{\theta}_{lt}'\} \\
= &\mathbb{P}_t^{\phi}\{\tilde{\phi}_t < \underline{\phi}_t'\} + \mathbb{P}_t^{\phi}\{\tilde{\phi}_t > \overline{\phi}_t'\} + \mathbb{P}_t^{\theta_l}\{\tilde{\theta}_{lt} < \underline{\theta}_{tl}'\} + \mathbb{P}_t^{\theta_l}\{\tilde{\theta}_{lt} > \overline{\theta}_{lt}'\} \\
\leq &\overline{\Phi}_t(\underline{\phi}_t') + 1 - \underline{\Phi}_t(\overline{\phi}_t') + \overline{\Theta}_t^l(\underline{\theta}_{lt}') + 1 - \underline{\Theta}_t^l(\overline{\theta}_{lt}') = \gamma.
\end{aligned} \tag{3-32}$$

Therefore, DRCC (3-26) is satisfied if the line overload does not happen when $(\tilde{\phi}_t, \tilde{\theta}_{lt})$ takes values at the vertices of polyhedron $\{(\tilde{\phi}_t, \tilde{\theta}_{lt}) | \tilde{\phi}_t \in [\underline{\phi}_t', \overline{\phi}_t']$ and $\tilde{\theta}_{lt} \in [\underline{\theta}_{lt}', \overline{\theta}_{lt}']\}$, which can be further written as deterministic constraint (3-33) by noticing $\underline{\theta}_{lt}' < \overline{\theta}_{lt}'$, i.e. $\forall l \in \mathcal{L}$:

$$\sum_{i\in\mathcal{G}} \pi_{gi}^l(p_{it}^g + \alpha_{it}^g \underline{\phi}_t') + \sum_{i\in\mathcal{S}} \pi_{si}^l(p_{it}^{sd} - p_{it}^{sc} + \alpha_{it}^s \underline{\phi}_t') - \sum_{i\in\mathcal{B}} \pi_{di}^l p_i^d - \underline{\theta}_{lt}' \leq C_l \tag{3-33a}$$

$$\sum_{i\in\mathcal{G}} \pi_{gi}^l(p_{it}^g + \alpha_{it}^g \overline{\phi}_t') + \sum_{i\in\mathcal{S}} \pi_{si}^l(p_{it}^{sd} - p_{it}^{sc} + \alpha_{it}^s \overline{\phi}_t') - \sum_{i\in\mathcal{B}} \pi_{di}^l p_i^d - \underline{\theta}_{lt}' \leq C_l \tag{3-33b}$$

$$-C_l \leq \sum_{i\in\mathcal{G}} \pi_{gi}^l(p_{it}^g + \alpha_{it}^g \underline{\phi}_t') + \sum_{i\in\mathcal{S}} \pi_{si}^l(p_{it}^{sd} - p_{it}^{sc} + \alpha_{it}^s \underline{\phi}_t') - \sum_{i\in\mathcal{B}} \pi_{di}^l p_i^d - \overline{\theta}_{lt}' \tag{3-33c}$$

$$-C_l \leq \sum_{i\in\mathcal{G}} \pi_{gi}^l(p_{it}^g + \alpha_{it}^g \overline{\phi}_t') + \sum_{i\in\mathcal{S}} \pi_{si}^l(p_{it}^{sd} - p_{it}^{sc} + \alpha_{it}^s \overline{\phi}_t') - \sum_{i\in\mathcal{B}} \pi_{di}^l p_i^d - \overline{\theta}_{lt}' \tag{3-33d}$$

As revealed in deterministic constraints (3-30), (3-31) and (3-33), the system can safely respond to random variable $\tilde{\phi}_t$ in the range of $[\underline{\phi}', \overline{\phi}']$ which we call the dispatchable range of total forecasting error. To ensure the system security, the system operator resorts to load shedding when $\tilde{\phi}_t$ exceeds $\overline{\phi}_t'$ and renewable curtailment when $\tilde{\phi}_t$ goes below $\underline{\phi}_t'$. Therefore, the reserve utilization costs (3-28) are the saturating linear function of total forecasting error $\tilde{\phi}_t$.

3.4.2 Evaluation of Worst-case Expectation

In the objective function (3-27), we need to evaluate the worst-case expectation of the reserve utilization costs which is a piece-wise linear function of the decision variables. The structure of the ambiguity set (3-5) allows the reformulation of the worst-case expectation as a LP, formally stated as

Lemma 3.2： Let $\hat{\phi}_t^{(1)}, \hat{\phi}_t^{(2)}, \cdots, \hat{\phi}_t^{(n)}$ be the ascendingly ordered samples of random variable $\tilde{\phi}_t$. Without loss of generality, assume $\hat{\phi}_t^{(k)} \leq 0, \forall k \leq m$ and $\hat{\phi}_t^{(k)} > 0, \forall k > m$. The ambiguity set $\mathcal{P}_t^{\phi}$ is constructed as in (3-5), i.e.

$$\mathcal{P}_t^{\phi} = \left\{\mathbb{P} \in \mathcal{P}_0([\underline{\phi}_t, \overline{\phi}_t]) \middle| \mathbb{P}\{\tilde{\phi}_t \leq \hat{\phi}_t^{(k)}\} \in [\underline{p}_t^k, \overline{p}_t^k], k = 1, \cdots, n\right\} \tag{3-34}$$

For notational convenience, let $\hat{\phi}_t^{(0)} = \underline{\phi}_t$, $\hat{\phi}_t^{(n+1)} = \overline{\phi}_t$ and $\underline{p}_t^{n+1} = \overline{p}_t^{n+1} = 1$. The worst-case expectation in (3-27) is equal to the optimum of the following LP by Lemma 2.4:

$$\begin{aligned}
&\max_{\mathbb{P}_t^\phi \in \mathcal{P}_t^\phi} \mathbb{E}_{\mathbb{P}_t^\phi}\{Q_t(\boldsymbol{\alpha}_t^g, \boldsymbol{\alpha}_t^s, \tilde{\phi}(t))\} \\
= &\min_{\substack{\underline{\lambda}_t^k, \overline{\lambda}_t^k \\ k=m-1,\cdots,m+2}} \sum_{k=m-1}^{m+2} (\overline{\lambda}_t^k \overline{p}_t^k - \underline{\lambda}_t^k \underline{p}_t^k) \\
&+ \sum_{k=1}^{m-2} (f_t^{dn}(\hat{\phi}_t^{(k-1)}) - f_t^{dn}(\hat{\phi}_t^{(k)}))\overline{p}_t^k \\
&+ \sum_{k=m+3}^{n} (f_t^{up}(\hat{\phi}_t^{(k)}) - f_t^{up}(\hat{\phi}_t^{(k+1)}))\underline{p}_t^k + f_t^{up}(\hat{\phi}_t^{(n+1)})\overline{p}_t^{n+1} \\
&\text{s.t.} \begin{cases}
\underline{\lambda}_t^k \geq 0,\ \overline{\lambda}_t^k \geq 0, k = m-1, \cdots, m+2 \\
\sum_{i=k}^{m+2}(\overline{\lambda}_t^i - \underline{\lambda}_t^i) + f_t^{up}(\hat{\phi}_t^{(m+3)}) \geq f_t^{dn}(\hat{\phi}_t^{(k-1)}), \\
\qquad\qquad k = m-1, m, m+1 \\
\sum_{i=k}^{m+2}(\overline{\lambda}_t^i - \underline{\lambda}_t^i) + f_t^{up}(\hat{\phi}_t^{(m+3)}) \geq f_t^{up}(\hat{\phi}_t^{(k)}), \\
\qquad\qquad k = m+1, m+2
\end{cases}
\end{aligned} \tag{3-35}$$

where

$$f_t^{up}(\tilde{\phi}_t) = g_t^{up}\tilde{\phi}_t - g_t^{up}(\tilde{\phi}_t - \overline{\phi}_t')^+ \tag{3-36}$$

$$f_t^{dn}(\tilde{\phi}_t) = -g_t^{dn}\tilde{\phi}_t - g_t^{dn}(-\tilde{\phi}_t + \underline{\phi}_t')^+ \tag{3-37}$$

$$g_t^{up} = \sum_{i\in\mathcal{G}} f_i^{g,up}\alpha_{it}^g + \sum_{i\in\mathcal{S}} f_i^{s,up}\alpha_{it}^s \tag{3-38}$$

$$g_t^{dn} = \sum_{i\in\mathcal{G}} f_i^{g,dn}\alpha_{it}^g + \sum_{i\in\mathcal{S}} f_i^{s,dn}\alpha_{it}^s. \tag{3-39}$$

□

3.4.3 Deterministic MILP Formulation

In summary, evaluating the worst-case expectation in (3-27) with the LP (3-35) and replacing the DRCC (3-24)∼(3-26) with the deterministic reformulation (3-30)∼(3-33), we can cast the proposed formulation into a **MILP model** for which off-the-shelf solvers are available.

3.4.4 Eliminating Inactive Line Capacity Constraints

The number of line capacity constraints (3-33) are $4 \times |\mathcal{L}| \times |\mathcal{T}|$ which could be prohibitively large for real-world power systems with small dispatch intervals. Fortunately, in practice, most of the line capacity constraints are inactive thus redundant for the optimization model. If the inactive constraints can be identified and eliminated before solving the problem,

the computational burden can be significantly reduced. Here we extend the fast identification method in [82] to the problem formulation in this chapter.

Consider the following problems:

$$\begin{aligned}&\Lambda^l_{t,max}(\tilde{\phi}_t,\tilde{\theta}_{lt})\left(\Lambda^l_{t,min}(\tilde{\phi}_t,\tilde{\theta}_{lt})\right)\\&=\arg\max_{\boldsymbol{p}^g_t,\boldsymbol{p}^s_t,\boldsymbol{\alpha}^g_t,\boldsymbol{\alpha}^s_t}(\min)\sum_{i\in\mathcal{G}}\pi^l_{gi}(p^g_{it}+\alpha^g_{it}\tilde{\phi}_t)+\sum_{i\in\mathcal{S}}\pi^l_{si}(p^s_{it}+\alpha^s_{it}\tilde{\phi}_t)-\sum_{i\in\mathcal{B}}\pi^l_{di}p^d_i-\tilde{\theta}_{lt}\end{aligned}\tag{3-40}$$

subject to

$$\sum_{i\in\mathcal{G}}(p^g_{it}+\alpha^g_{it}\tilde{\phi}_t)+\sum_{i\in\mathcal{S}}(p^s_{it}+\alpha^s_{it}\tilde{\phi}_t)=\sum_{i\in\mathcal{B}}p^d_i+\tilde{\phi}_t \tag{3-41a}$$

$$\underline{P}^g_i\le p^g_{it}+\alpha^g_{it}\underline{\phi}'_t \tag{3-41b}$$

$$p^g_{it}+\alpha^g_{it}\overline{\phi}'_t\le\overline{P}^g_i \tag{3-41c}$$

$$-\overline{P}^{sc}_i\le p^s_{it}+\alpha^s_{it}\underline{\phi}'_t \tag{3-41d}$$

$$p^s_{it}+\alpha^s_{it}\overline{\phi}'_t\le\overline{P}^{sd}_i. \tag{3-41e}$$

where (3-41a) is obtained by multiplying (3-13) by $\tilde{\phi}_t$ and adding to (3-12); (3-41b)$\sim$(3-41c) are deduced from (3-15), (3-30) and (3-31). Therefore, the feasible sets of the above optimization problems are relaxations of the feasible set of the original MILP model. Minimization (maximization) w.r.t. the feasible set defined by (3-41) yields a lower (upper) bound of the minimum (maximum) w.r.t the feasible set of the original MILP model. The objective function (3-40) is just the possible line flow at each line in each time period. Similar to the analysis in [82, 100], we have the following lemma.

Lemma 3.3： For any $l\in\mathcal{L}$ and $t\in\mathcal{T}$, we have

- If $\Lambda^l_{t,max}(\underline{\phi}_t,\underline{\theta}^l_t)\le C_l$, constraint (3-33a) is inactive;
- If $\Lambda^l_{t,max}(\overline{\phi}_t,\underline{\theta}^l_t)\le C_l$, constraint (3-33b) is inactive;
- If $\Lambda^l_{t,min}(\underline{\phi}_t,\overline{\theta}^l_t)\ge -C_l$, constraint (3-33c) is inactive;
- If $\Lambda^l_{t,min}(\overline{\phi}_t,\overline{\theta}^l_t)\ge -C_l$, constraint (3-33d) is inactive. □

To simplify the above LP (3-40)(3-41), we further define $\mathcal{H}=\mathcal{G}\cup\mathcal{S}$ and for any $i\in\mathcal{H}$:

$$p_{it}=\begin{cases}p^g_{it}+\alpha^g_{it}\tilde{\phi}_t-\underline{P}^g_i,\ i\in\mathcal{G}\\p^s_{it}+\alpha^s_{it}\tilde{\phi}_t+\overline{P}^c_{si},\ i\in\mathcal{S}\end{cases}\tag{3-42a}$$

$$\pi^l_i=\begin{cases}\pi^l_{gi},\ i\in\mathcal{G}\\\pi^l_{si},\ i\in\mathcal{S}\end{cases}\tag{3-42b}$$

$$\overline{P}_i=\begin{cases}\overline{P}^g_i-\underline{P}^g_i,\ i\in\mathcal{G}\\\overline{P}^{sd}_i+\overline{P}^{sc}_i,\ i\in\mathcal{S}\end{cases}\tag{3-42c}$$

$$D_t=\sum_{i\in\mathcal{B}}\hat{p}^d_{it}-\sum_{i\in\mathcal{G}}\underline{P}^g_i+\sum_{i\in\mathcal{S}}\overline{P}^{sc}_i\tag{3-42d}$$

$$K_t^l = -\sum_{i\in\mathcal{B}} \pi_{di}^l \hat{p}_{it}^d + \sum_{i\in\mathcal{G}} \pi_{gi}^l \underline{P}_i^g - \sum_{i\in\mathcal{S}} \pi_{si}^l \overline{P}_i^{sc} \tag{3-42e}$$

Then we have

$$\begin{aligned} &\Lambda_{t,max}^l(\tilde{\phi}_t, \tilde{\theta}_{lt}) \left(\Lambda_{t,min}^l(\tilde{\phi}_t, \tilde{\theta}_{lt})\right) \\ &= \arg\max_{\boldsymbol{p}_t}(\min) \sum_{i\in\mathcal{H}} \pi_i^l p_{it} + K_t^l - \tilde{\theta}_{lt} \\ &\text{s.t.} \begin{cases} \sum_{i\in\mathcal{H}} p_{it} = D_t + \tilde{\phi}_t \\ 0 \le p_{it} \le \overline{P}_i \end{cases} \end{aligned} \tag{3-43}$$

LP (3-43) has a analytical solution according to the analysis in [82]. Let $i_1, i_2, \cdots, i_{|\mathcal{H}|}$ be a permutation of $1, 2, \cdots, |\mathcal{H}|$ such that $\{\pi_{i_1}^l, \pi_{i_2}^l, \cdots, \pi_{i_{|\mathcal{H}|}}^l\}$ are in descending (ascending) order, and there exists an integer $1 \le m \le |\mathcal{H}|$ such that $\sum_{k=1}^{m-1} \overline{P}_{i_k} \le D_t + \tilde{\phi}_t \le \sum_{k=1}^{m} \overline{P}_{i_k}$. Then

$$\begin{aligned} &\Lambda_{t,max}^l(\tilde{\phi}_t, \tilde{\theta}_{lt}) \left(\Lambda_{t,min}^l(\tilde{\phi}_t, \tilde{\theta}_{lt})\right) \\ &= \sum_{k=1}^{m-1} (\pi_{i_k}^l - \pi_{i_m}^l)\overline{P}_{i_k} + \pi_{i_m}^l (D_t + \tilde{\phi}_t) + K_t^l - \tilde{\theta}_{lt}. \end{aligned} \tag{3-44}$$

Based on the analytical expression (3-44), lemma 3.3 gives a computationally cheap way to identify most of the inactive line capacity constraints in (3-33).

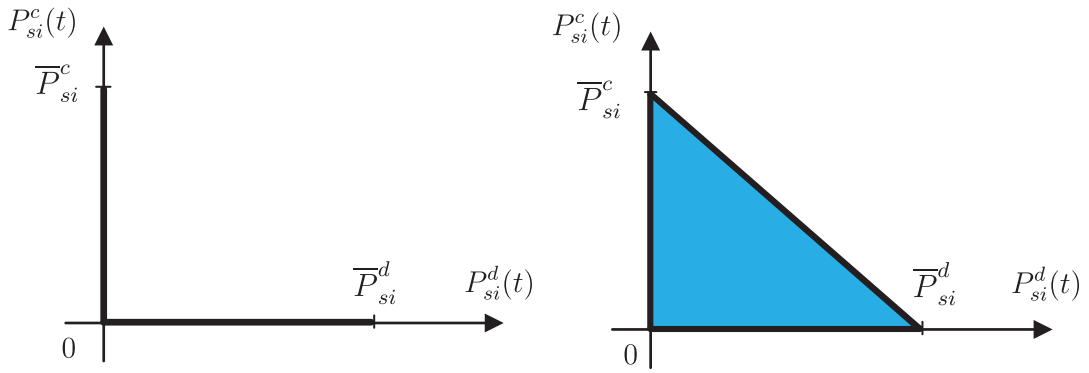
3.4.5 Convex Relaxation

The non-convexity of the proposed optimization model comes only from the binary variables introduced in (3-18) to avoid simultaneous charging and discharging of energy storages. The analysis in [101] and [102] show that simultaneous charging and discharging could only happen when the local marginal price goes below a negative threshold value, which is very unusual in practical operation. Considering such feature, we employ an iterative scheme similar to the successive constraint enforcement in [74, 75]. In our implementation, we first relax all the constraint (3-18) to its convex hull:

$$p_{it}^{sc} \ge 0,\ p_{it}^{sd} \ge 0 \tag{3-45a}$$

$$p_{it}^{sc}/\overline{P}_i^{sc} + p_{it}^{sd}/\overline{P}_i^{sd} \le 1 \tag{3-45b}$$

which is illustrated in Fig. 3-1. In this way, we obtain a LP relaxation to the original MILP. If the solution of relaxed model happens to lie in the feasible set of the original MILP model, we can conclude that this solution is also the global optimal solution of the original MILP problem. Therefore, after solving the LP model, we check whether $p_{it}^{sc} p_{it}^{sd} = 0$ for all $i \in \mathcal{S}, t \in \mathcal{T}$. If so, the solution to the orginal MILP model is found. Otherwise, the constraint (3-45) has to be changed back to constraint (3-18) for those i and t where $p_{it}^{sc} p_{it}^{sd} > 0$. Hence we obtain a



(a) Feasible set of constraint (3-18) (b) Convex hull defined by (3-45)

Figure 3-1 Illustration of Convex Relaxation of (3-18) to (3-45).

relaxed MILP model which is still much simpler than the orignal MILP model. Again we solve the relaxed model and then check the exactness of the relaxation. This process is repeated until $p^{sc}_{it}p^{sd}_{it} = 0$ for all $i \in \mathcal{S}, t \in \mathcal{T}$.

3.5 Numerical Results

The proposed formulation and method were programmed in MATLAB with Gurobi as the MILP and LP solver running on a Win 8 PC with a 3.0GHz CPU and 24 GB RAM. The simulation was carried out on IEEE-118 bus system modified according to [103]. Five wind farms were installed at bus 16, 37, 48, 75 and 83 with each capacity of 100MW. Storage was assumed to be installed at each non-generator bus. The energy capacity for each storage was 32 MW, and the charge & discharge power capacity were both set to be 8 MW/h. The charge & discharge efficiency were both 0.9. The ramp rates over each hour for conventional generator were set to be 50% of the rated capacity. We considered a time horizon of 24 h with each time step 30 min. The half-hourly forecasting load and wind generation profile were obtained from [75] and NREL WIND Toolkit. All uncertainties were assumed to originate from the forecasting errors of wind generation. Different types of probability distributions were employed to generate "realistic" data of wind power forecasting errors whose mean and variance were set according to the typical day-ahead forecasting errors in U.S. reported in [85]. The confidence level α for the confidence bands of CDF is set to 0.05. In addition, the reserve availability and utilization prices of generators and storages were randomly selected from the ranges shown in Table 3-1.

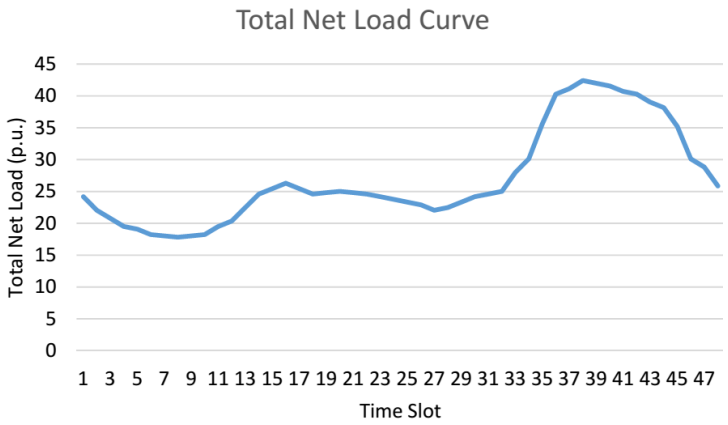
Table 3-1 Reserve Availability and Utilization Prices

$R^{g,up}_i/R^{g,dn}_i$	$f^{g,up}_i/f^{g,dn}_i$	$R^{s,up}_i/R^{s,dn}_i$	$f^{s,up}_i/f^{s,dn}_i$	c^{sd}_i/c^{sc}_i
0.1∼0.3 c^g_i	0.8∼1.1 c^g_i	0.1∼0.3 c^{sd}_i/c^{sc}_i	0.8∼1.1 c^{sd}_i/c^{sc}_i	15/10 \$/WMh

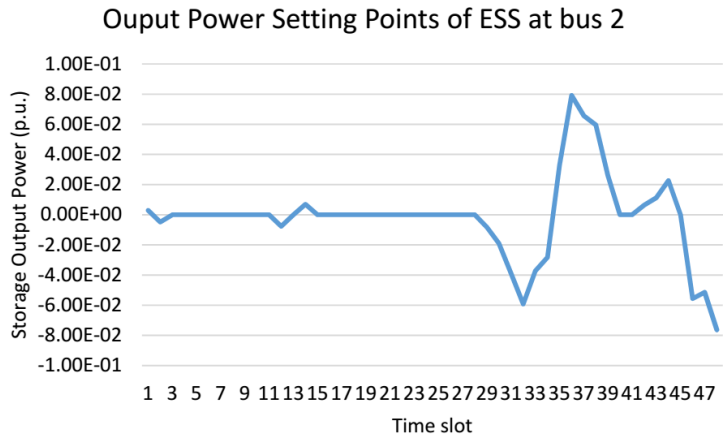
The proposed method is data-driven and distribution-free, so it can deal with wind power forecasting errors following any probability distributions. Thus, we tested the method with wind power forecasting errors generated from normal, laplace, beta and hyperbolic distributions, and Monte Carlo simulations with 10^6 samples were employed to assess the performance of the proposed method. Fig. 3-8 illustrates the evolution of optimization objective function (solid blue line) and the operation costs from Monte Carlo simulation (dotted red line) as the number of available historical data increases. Firstly, we observe that the operation costs from Monte Carlo simulation are always upper bounded by the objective function of the optimization model. This exhibits the distributional robustness of the proposed method: the objective function represents the costs w.r.t the worst-case distribution in the ambiguity set whereas the underlying true distribution could be different from the worst-case one. Secondly, as more historical data is available, both the objective functions and the simulated costs decrease, which shows the value of data: the more historical data is employed, the less conservative the solution is. Finally, the gaps between the objective functions and the simulated costs are narrowed by incorporating more historical data which reveals that the ambiguity set shrinks to the underlying true probability distribution as the number of historical data increases. Table 3-2 further lists the percentage gaps between objective functions and simulated costs. Figure 3-2 shows the forecasted gross load curve and the corresponding optimal setting point curve of ESS and generators when 10^4 data is available.

Since the proposed problem formulation is a direct extension of the robust multi-period OPF [75], we compared the proposed CDF-based DRO approach with the RO approach in [75]. The RO approach can be implemented in our problem formulation by: 1) replacing the DRCC (3-24)$\sim$(3-26) with linear robust constraints using the support of the random variable as the uncertainty set; 2) eliminating the reserve utilization term in the objective function (3-27). The operation costs of the strategies obtained by both methods are compared in Fig. 3-3. In Fig. 3-3(a), the wind power forecasting follows beta distribution with different variances. The proposed CDF-based DRO approach achieves lower operation costs than RO approach under each level of forecasting errors. The difference of the operation costs tends to increase as the variances of the forecasting error increase. In Fig. 3-3(b), the similar comparison is carried out under different types of distributions of the forecasting errors. It is also shown that the proposed CDF-based DRO approach obtains more economical operation strategies than the RO approach. In short, the proposed CDF-based DRO approach captures the detailed probabilistic information from historical data while the RO approach ignores such information. When more data is available, the proposed method can produce a more economic strategy whereas the RO does not have a mechanism to take advantage of more data.

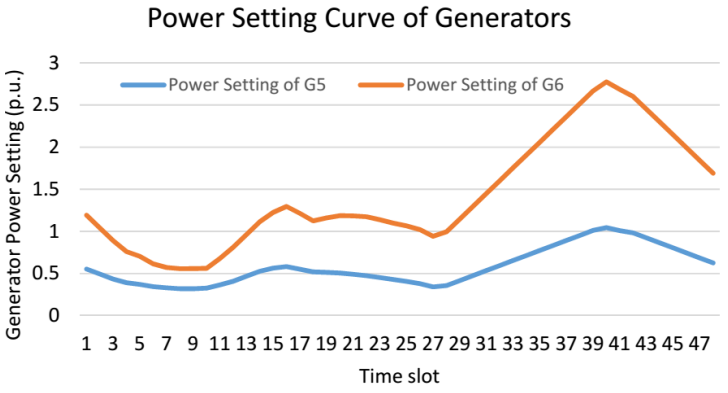
Another major competitor of the CDF-based DRO is the moment-based DRO appears in [38, 42, 63, 64], etc. The ambiguity set for moment-based DRO is the set of all probability



(a) Total net load curve



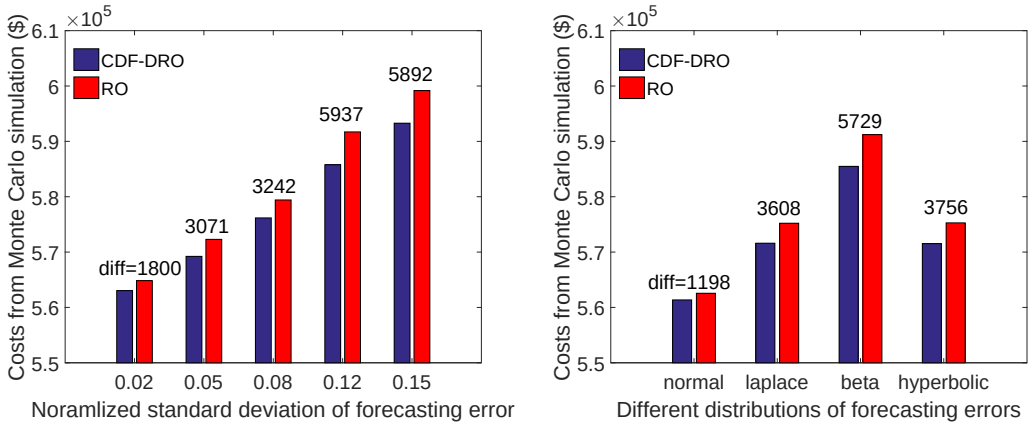
(b) Output power setting point curve of the storage at bus 2



(c) Output power setting point curve of G5 and G6

Figure 3-2 Cost comparison between CDF-based DRO and RO

distributions with given mean and covariance (usually sample mean and sample covariance). To initiate a meaningful comparison, we implement the moment-based DRO in our problem formulation as follows: 1) the chance constraints (3-24)∼(3-26) can be reformulated as SOCP constraints by leveraging the Chebyshev inequality $\mathbb{P}\{|\tilde{\xi} - \mu| \geq \sqrt{1/\rho}\sigma\} \leq \rho$ [42]; 2) the evaluation of worst-case expectation in (3-27e) can be also cast into a SOCP using duality



(a) Different standard deviations

(b) Cost comparison between CDF-based DRO and RO

Figure 3-3 Cost comparison between CDF-based DRO and RO

theory of moment problem. Fig. 3-4 compares the operational costs by two methods when different number of historical data is available. The curves under different types of underlying true distributions show the similar pattern. The costs of moment-based DRO is always higher than that of the CDF-based DRO and the difference is enlarged when more data is at hand. By merely relying on the information of the first two moments, the moment-base DRO is unable to fully take advantage of the abundance of data, which is in stark contrast to the data-exploiting feature of the CDF-based DRO. Fig. 3-5 and fig. 3-6 confirm the foregoing conclusion by observing the violation probability of chance constraint (3-24) and (3-25). In this group of test, we have set the $\beta_1 = \beta_2 = 0.05$. Fig. 3-5 and Fig. 3-6 show that both methods ensure higher reliability level than required due to their "distributionally robust" nature. As more data is available, the proposed CDF-based DRO gradually and safely reduce the guaranteed reliability level to pursue higher economic efficiency. In contrast, the conservatism of the moment-based DRO remains significant even with 10^6 data. To further reveal the nature of both methods, we then investigate the relationship between β and the minimal τ that ensure $\mathbb{P}[|\xi - \mu| \leq \tau] \geq 1 - \beta, \forall \mathbb{P} \in \mathcal{P}$ when $\mathcal{P}$ is the moment-based ambiguity set or CDF-based ambiguity set constructed from the different number of data. It is shown in Fig. 3-7 that the true relation between β and τ for the specific distribution under study is always upper bounded by those provided by the DRO approaches. The curve given by the moment-based DRO is very far away from the true curve, and the curves provided by the CDF-based DRO approach the true curve as more and more data is available. In summary, the CDF-based DRO extracts much more probabilistic information from data than the moment-based DRO approach, which leads to less conservative operation strategy. In addition, when more data is at hand, the moment-based DRO can merely have a more accurate guess of the mean and covariance, whereas the proposed CDF-based DRO can extract more detailed information about the whole distribution.

The CDF-based DRO thus has much stronger ability to exploit data.

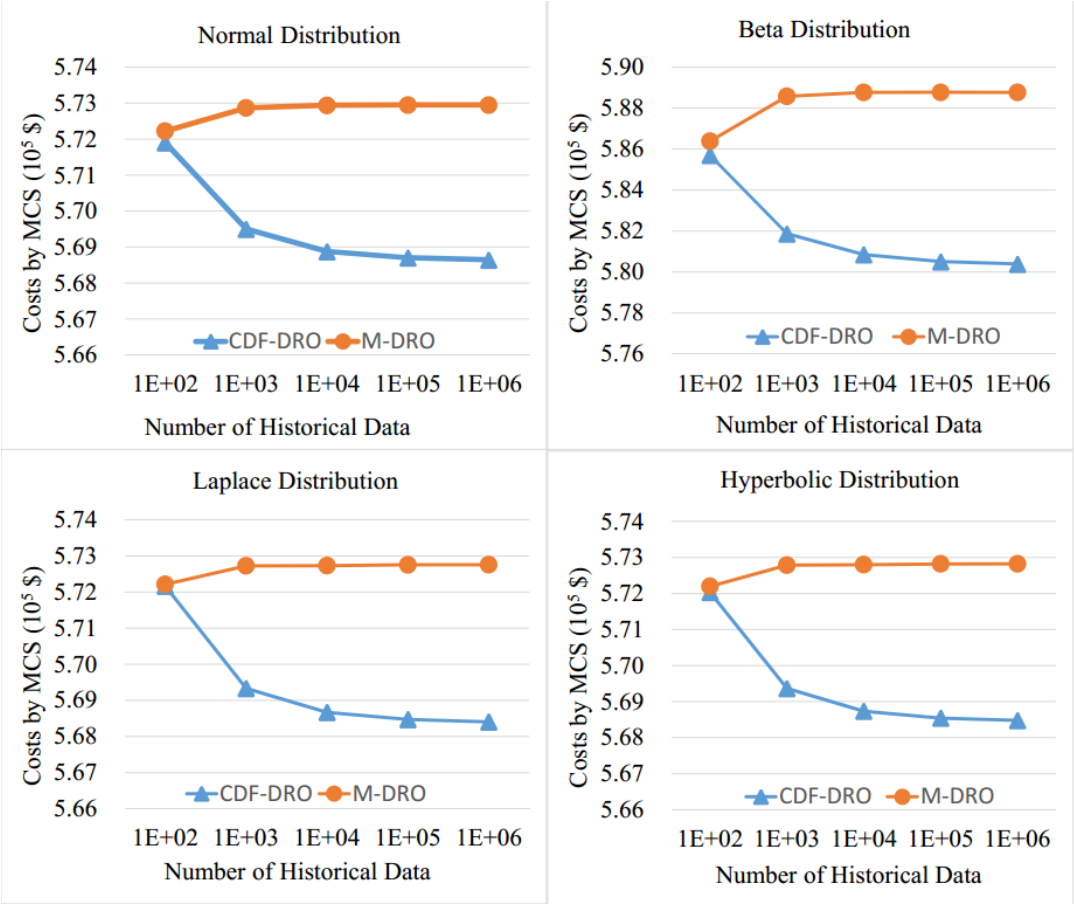


Figure 3-4 Comparison between the CDF-based DRO and the moment-based DRO for the operational costs by MCS under different types of distributions.

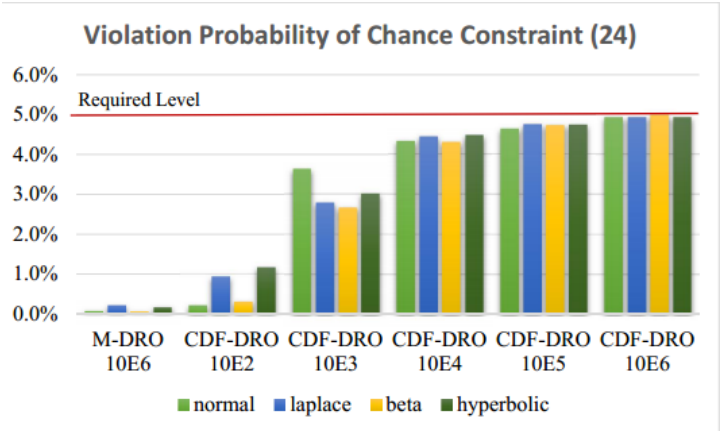


Figure 3-5 Comparion of the violation probability of chance constraint (24) between the moment-based DRO and the CDF-based DRO with different number of data.

The focus is then given to the computational efficiency of the proposed method. Table 3-3 lists the solver time of the proposed CDF-based DRO approach using the different number of data points along with the solver time of the RO and the moment-based DRO approaches. The

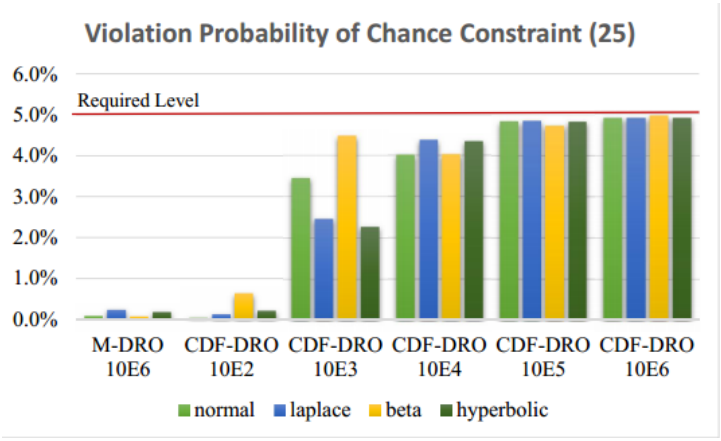


Figure 3-6 Comparion of the violation probability of chance constraint (25) between the moment-based DRO and the CDF-based DRO with different number of data.

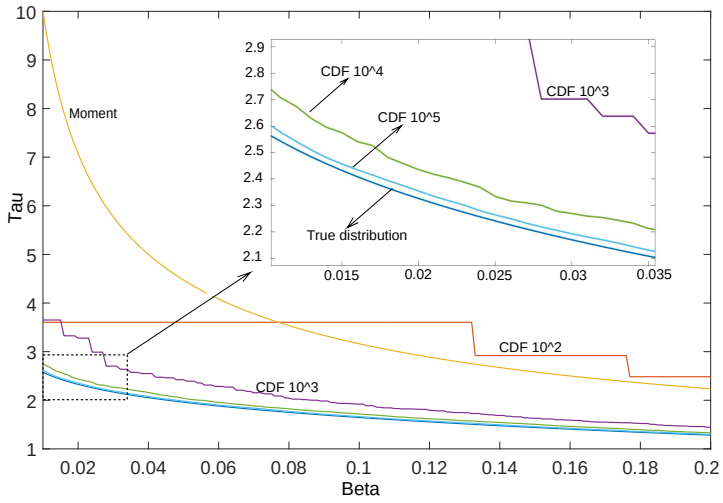
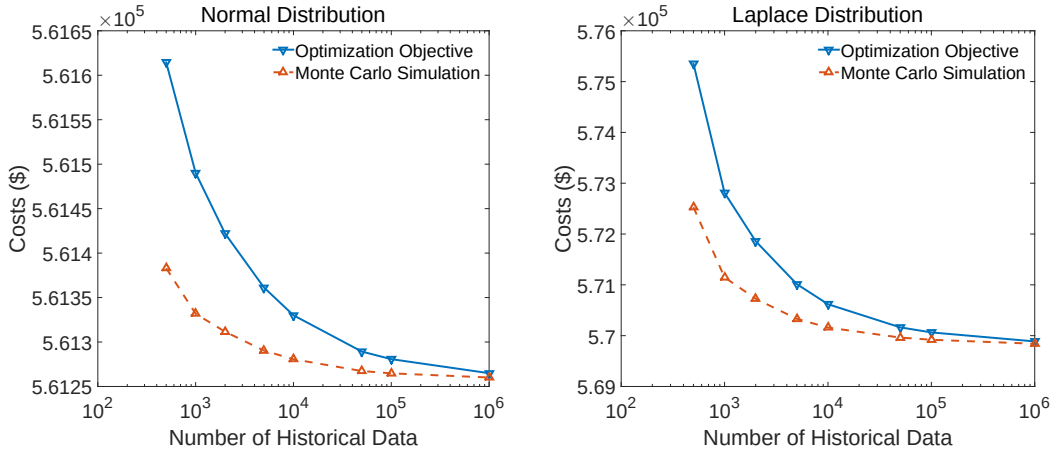


Figure 3-7 The relationship between β and the minimal τ which ensure $\mathbb{P}[|\xi-\mu| \leq \tau] \geq 1-\beta, \forall \mathbb{P} \in \mathcal{P}$ when $\mathcal{P}$ is the moment-based ambiguity set or CDF-based ambiguity set constructed from different number of data.

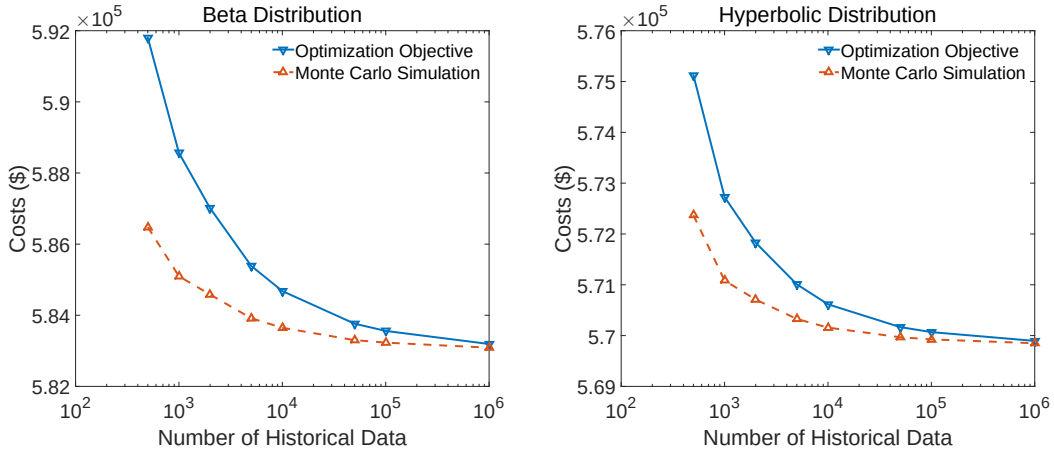
solver time of the proposed method does not necessarily grow with the number of historical data due to the fact that the scale of the proposed MILP model is irreverent to the number of data points. The variation of the solver time with the number of data points is just due to numerical issues of the solver. Compared with the RO approach, the computational burden of the proposed CDF-based DRO approach is slightly more intensive on average due to the larger number of decision variables and constraints. The moment-based DRO, on the other hand, can be solved much slower than the other two approaches due to its SOCP formulation.

Table 3-4 demonstrates the effectiveness of the inactive constraint elimination and convex relaxation procedures discussed in section 3.4. It is shown that more than 88% of the line capacity constraints are identified to be inactive thus redundant to the optimization model. In all our tests, the average solver time of the proposed method without inactive constraint elimination is around 38.7s. After elimination, the average solver time reduces to 9.6s which is

approximately 1/4 of that without redundant constraint elimination. Moreover, the relaxation of constraint (3-18) to constraint (3-45) is 100% exact for the energy storages, which indicates the local marginal prices at the storage buses are always beyond the threshold values for exact relaxation [101, 102]. The underlying physical reason for this phenomenon is the relatively adequate transmission and storage capacity of the case under study [102]. Both the elimination of the inactive line capacity constraints and the convex relaxation of constraint (3-18) significantly contribute to the improvement of numerical tractability.



(a) Normal (b) Laplace



(c) Beta (d) Hyperbolic

Figure 3-8 Evolution of objective function and simulated costs as the increase of available historical data.

3.6 Conclusion

In this chapter, an energy-reserve-storage co-optimization model and a data-driven distributionally robust method are proposed to achieve economical and reliable operation of power systems with variable and uncertain renewable sources. Compared with the SP approach, the proposed method assumes no prior knowledge of the probability distribution of the uncertain-

Table 3-2 Percentage Gap Between the Objective Function and the Costs by Monte Carlo Simulation

data num.	500	1000	2000	5000	10000	50000
normal	0.041%	0.028%	0.020%	0.013%	0.009%	0.004%
laplace	0.494%	0.291%	0.198%	0.120%	0.081%	0.035%
beta	0.907%	0.593%	0.416%	0.252%	0.177%	0.079%
hyperbolic	0.481%	0.288%	0.196%	0.119%	0.080%	0.035%

Table 3-3 Solver Time (sec.) of the RO, the moment-based DRO and the CDF-based DRO.

method	RO	M-DRO	CDF-DRO (10^2)	CDF-DRO (10^3)	CDF-DRO (10^4)	CDF-DRO (10^5)
normal	3.8	146.3	6.1	3.6	7.9	11.6
laplace	4.4	132.9	6.2	4.3	7.8	12.3
beta	6.7	187.6	6.2	8.8	10.7	11.7
hyperbolic	3.9	128.5	7.1	3.8	4.3	12.5

Table 3-4 Percentages of Identified Inactive Line Capacity Constraints and Exact Convex Relaxations

distribution	normal	laplace	beta	hyperbolic
inactive line constraint (%)	89.57%	88.78%	88.55%	88.82%
relaxation exactness (%)	100%	100%	100%	100%

ties and achieve operational robustness by considering the worst-case distribution consistent with the observed data. Compared with the RO approach, the proposed method exploits detailed probabilistic information learned from historical data and the conservatism of the solution can be reduced by incorporating more historical data. Numerical studies demonstrate the favorable features of the proposed methods.

4 Distributionally Robust Chance-Constrained AC-OPF with Wasserstein Metric

4.1 Introduction

Large-scale VRE sources have been integrated into modern power systems with ever-increasing penetration. The uncertainties brought by VREs are no longer negligible and pose considerable risk to power system security. Risk management has been identified as one of the major challenges of integrating high penetrations of renewable energy [104]. Therefore, developing optimization models and solution approaches with risk consideration is crucial for the reliable and economical operation of power systems. In recent years, we have seen extensive literature dealing with optimization of power system operation under uncertainties.

Stochastic programming (SP) [21, 22], robust optimization (RO) [75, 86, 97, 105] and distributionally robust optimization (DRO) [35–42, 106] have been employed to tackle uncertainties in power system operation. SP assumes the uncertainties follow a known probability distribution and transforms the stochastic problems into deterministic ones either by sampling or by analytical reformulation. On the contrary, RO does not require any probabilistic information of the uncertainties. A deterministic uncertainty set is constructed to include all possible realizations of the random variable. RO seeks strategies that perform best w.r.t. the worst-case realization in the uncertainty set. In fact, the uncertainties do obey some underlying probability distribution, but this distribution is not known as a priori and only some sample data is available. In other words, the underlying true distribution is ambiguous to the decision makers. To remedy the SP's specificity and RO's ignorance of the probabilistic information, DRO assumes that the true distribution lies in an ambiguity set and immunizes the operation strategies against all distributions in the ambiguity set. The most popular ambiguity set is moment-based, i.e. the set of all probability distributions with given mean and covariance [37–40, 42]. However, only the first two moments do not constitute a detailed characterization of the true distribution. Especially when we have a large amount of data at hand, much more probabilistic information can be extracted and exploited not just the first two moments. Intuitively, the more data is available, the more we know about the true distribution. Therefore, a desirable ambiguity set should be made smaller by incorporating more data. Moment-based ambiguity sets, obviously, do not possess such feature.

In the framework of SP or DRO, the chance constraints, which require the security constraints hold with a specified probability level, have also been introduced to power system operation problems, especially the OPF problem [41–43, 107, 108]. In [41, 43, 107], by assuming the uncertainties follow Gaussian or Student's t distribution, the chance constraints

are cast into second-order cone programming (SOCP) constraints, and the uncertainties of the mean and covariance are further considered to achieve distributional robustness. In [42, 108], no assumption about the distribution type is used. The chance constraints are required to be satisfied for any probability distribution in the moment-based ambiguity set. All above existing works employ the DC power flow model without consideration for the voltage magnitudes and reactive power, which may lead to unpractical and unsafe operation strategy. Operated under the strategy obtained with DC power flow model, the system could be exposed to great risk of under- or over-voltage at some critical buses, and the generators would be forced to violate the excitation limits. Therefore, extending the chance-constrained OPF framework to include AC power flow model is a crucial step toward a better operation strategy under uncertainties.

In this chapter, we propose a Wasserstein-metric-based distributionally robust chance-constrained approximate AC-OPF. The contributions are as follows.

1. In the OPF formulation, to overcome the inaccuracy and V/Q unconcern of DC power flow model, we develop a novel approximate AC power flow model which is a combination of an exact nonlinear AC model at the nominal operation point and an approximate linear model to represent the system response under uncertainties. It largely inherits the accuracy of the full AC model and also maintains the tractability of linear power flow model for use in stochastic optimization.
2. We apply the recent results [109, 110] of data-driven DRO with Wasserstein metric to the approximate AC-OPF model. To our knowledge, this is the first time the Wasserstein-metric-based ambiguity set is introduced to the OPF problems, which provides a promising alternative for the moment-based ambiguity set. The method is data-driven without any presumption on the probability distribution of the uncertainties. With only a limited number of data, the chance constraints w.r.t. the underlying true probability distribution can be robustly guaranteed. The more historical data is available, the less conservative the solution is.
3. Note that the naive applications of the DRO approach discussed in [109, 110] would lead to very unscalable implementations, i.e. the computational burden grows heavily with the number of data and dimension of uncertainties. To overcome such drawbacks, beyond a direct application of the general theories, the specific problem structures of the proposed OPF formulation are fully exploited to develop a scalable and efficient solution method.

Nomenclature

Boldface lower-case letter $\boldsymbol{x}$ represents a real vector and its i-th element is denoted by x_i. Boldface upper-case letter $\boldsymbol{A}$ represents a matrix with its $(i,j)^{\text{th}}$ element denoted by A_{ij}. Random vectors are written as boldface lower-case letter with tildes, i.e. $\tilde{\boldsymbol{y}}$. Given an index set $\mathcal{S}$, the subvector of $\boldsymbol{x}$ indexed by $\mathcal{S}$ is denoted by $\boldsymbol{x}_{\mathcal{S}}$. The imaginary unit is denoted by j. We

use parentheses to construct vectors from comma separated lists as $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}) = [\boldsymbol{x}^T, \boldsymbol{y}^T, \boldsymbol{z}^T]^T$. In addition, the following special symbols are used in our problem formulation and derivation:

n_b, n_l	Number of buses and lines.
$\mathcal{R}$, $\mathcal{S}$, $\mathcal{L}$	Index sets for reference bus, PV buses and PQ buses, respectively.
$\boldsymbol{\theta}, \boldsymbol{v}$	$n_b \times 1$ vectors of nominal bus voltage angles and magnitudes.
$\boldsymbol{f}$	$n_l \times 1$ vector of nominal line MW flow.
$\boldsymbol{Y}, \boldsymbol{G}, \boldsymbol{B}$	$\boldsymbol{Y} = \boldsymbol{G} + j\boldsymbol{B}$ is the system admittance matrix.
$\boldsymbol{B}'$	The susceptance matrix without shunt elements.
$\boldsymbol{G}^l$, $\boldsymbol{B}^l$	$n_l \times n_b$ matrices with $G^l_{ki} = -G^l_{kj} = G_{ij}$, $B^l_{ki} = -B^l_{kj} = B_{ij}$, and other elements being zeros.
$\boldsymbol{p}^g, \boldsymbol{q}^g$	$n_b \times 1$ vectors of nominal active and reactive power injection from generators. $p^g_i = q^g_i = 0$ if no generator at bus i.
$\overline{\boldsymbol{p}}^g, \underline{\boldsymbol{p}}^g$	$n_b \times 1$ vectors of upper and lower output power limits of generators. $\overline{p}^g_i = \underline{p}^g_i = 0$ if no generator at bus i.
$\boldsymbol{\alpha}$	$n_b \times 1$ vectors of AGC participation factors of generating units. $\alpha_i = 0$ if no generator at bus i.
$\overline{\boldsymbol{r}}, \underline{\boldsymbol{r}}$	$n_b \times 1$ vectors of upward and downward regulating reserves of generating units. $\overline{r}_i = \underline{r}_i = 0$ if no generator at bus i.
$\overline{\boldsymbol{c}}, \underline{\boldsymbol{c}}$	$n_b \times 1$ vectors of upward and downward regulating reserves prices.
$\boldsymbol{p}^w, \boldsymbol{q}^w$	$n_b \times 1$ vectors of nominal active and reactive power injection from VRE.
$\boldsymbol{p}^l, \boldsymbol{q}^l$	$n_b \times 1$ vectors of nominal active and reactive power consumption of load.
$\tilde{\boldsymbol{\xi}}$	$n_b \times 1$ random vectors representing forecasting errors of VRE generation.
$\mathbb{P}$	A probability distribution (measure).
$\mathbb{E}_{\mathbb{P}}$	Expectation with probability distribution $\mathbb{P}$.
$\mathcal{P}(\Xi)$	Set of all distributions with support Ξ.
$(x)^+$	$\max\{x, 0\}$.
$\boldsymbol{A} \circ \boldsymbol{B}$	Hadamard product of matrix $\boldsymbol{A}$ and $\boldsymbol{B}$, i.e. $C_{ij} = A_{ij}B_{ij}$ if $\boldsymbol{C} = \boldsymbol{A} \circ \boldsymbol{B}$.

4.2 Problem Formulation

4.2.1 Power Flow and Its Control under Unceratainties

In a VRE integrated power system, the active and reactive power balance is governed by the following equations:

$$\boldsymbol{p}^g + \boldsymbol{p}^w - \boldsymbol{p}^l = \mathscr{P}(\boldsymbol{\theta}, \boldsymbol{v}) \tag{4-1a}$$

$$\boldsymbol{q}^g + \boldsymbol{q}^w - \boldsymbol{q}^l = \mathscr{Q}(\boldsymbol{\theta}, \boldsymbol{v}) \tag{4-1b}$$

where

$$\mathscr{P}(\boldsymbol{\theta}, \boldsymbol{v}) = (\boldsymbol{G} \circ \boldsymbol{W} \circ \cos\boldsymbol{\Theta} + \boldsymbol{B} \circ \boldsymbol{W} \circ \sin\boldsymbol{\Theta})\, \mathbf{1} \tag{4-2a}$$

$$\mathscr{Q}(\boldsymbol{\theta}, \boldsymbol{v}) = (-\boldsymbol{B} \circ \boldsymbol{W} \circ \cos\boldsymbol{\Theta} + \boldsymbol{G} \circ \boldsymbol{W} \circ \sin\boldsymbol{\Theta})\, \mathbf{1} \tag{4-2b}$$

with $\boldsymbol{W} = \boldsymbol{v}\boldsymbol{v}^\top$ and $\boldsymbol{\Theta} = \boldsymbol{\theta}\mathbf{1}^\top - \mathbf{1}\boldsymbol{\theta}^\top$. For any fixed amount of renewable energy generation $\boldsymbol{p}^w + j\boldsymbol{q}^w$ and load demand $\boldsymbol{p}^l + j\boldsymbol{q}^l$, the conventional OPF addresses the problem of designing the optimal operation point $(\boldsymbol{\theta}, \boldsymbol{v}, \boldsymbol{p}^g, \boldsymbol{q}^g)$ which minimizes the generation costs while satisfying the system balance and security constraints. With increasing level of VRE penetration, the power injections across the network are facing continuous fluctuations due to the variability and unceratainty of VRE. Under such circumstance, a deterministic optimal operation point is no longer enough to guide the system operation. Therefore, the control mechanism to couple with renewable and load unceratinties must be taken into account in the OPF formulation. The most widely used control systems in response to volatility are the AGC and AVR shown in Fig. 4-1 and Fig. 4-2 [111]. The AGC is a centralized control scheme to maintain the real-time active power balance by distributing the system power imbalance to each generating unit according to the corresponding partication factor. The AVR, on the other hand, is a decentralized control scheme to keep fixed voltage magnitudes at the buses equipped with reactive power scources. The most classic AVRs are those implemented in the generator exicitation systems, shown in Fig. 4-2, which maintain the stator voltages at the set values. With the implementation of the AGC and AVRs, the trasition of system operation point when renewable generation deviating from forecasting values is completely determined by the power flow equations (4-1). The new operation point can be computed by solving the conventional power flow problem with one generator bus as the reference bus, other generator buses as PV buses and the rest buses as PQ buses. Therefore, the new OPF formulation thus targets at deciding the optimal nominal operation point $(\boldsymbol{\theta}, \boldsymbol{v}, \boldsymbol{p}^g, \boldsymbol{q}^g)$ as well as the participation factors $\boldsymbol{\alpha}$ which provide a statistically economic and reliable system performance under uncertainties. However, the system response under uncertainties as described above is an inexplicit function given by the nonlinear power flow equations (4-1), which poses great challenge for formulating a tractable stochastic optimization problem. To circumvent this difficulty, an explicit surrogate formula is needed to approximate the exact implicit system response.

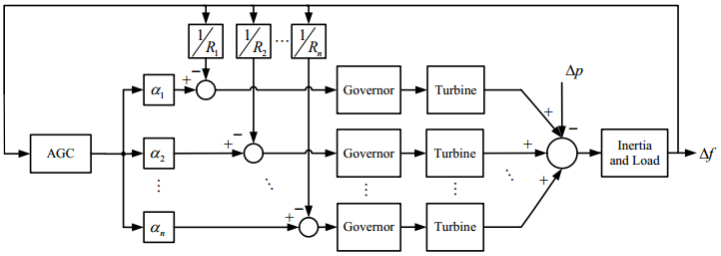


Figure 4-1 Automatic Generation Control

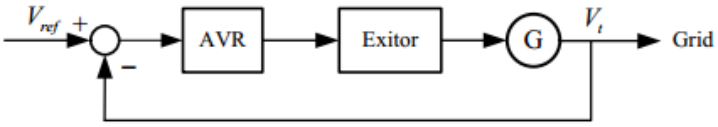


Figure 4-2 Automatic Voltage Regulation

4.2.2 Linear Power FLow Response Model under Uncertainties

Recent advances in linear power flow model have opened a way for developing simple explicit formula for system response under uncertainties. The following linear power flow (LPF) model was developed in [112]:

$$\begin{bmatrix} \boldsymbol{p} \\ \boldsymbol{q} \end{bmatrix} = - \begin{bmatrix} \boldsymbol{B}' & -\boldsymbol{G} \\ \boldsymbol{G} & \boldsymbol{B} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{v} \end{bmatrix} \tag{4-3}$$

$$\boldsymbol{f} = \boldsymbol{G}^l \boldsymbol{v} - \boldsymbol{B}^l \boldsymbol{\theta} \tag{4-4}$$

where $\boldsymbol{p}$ and $\boldsymbol{q}$ are the $n_b \times 1$ vectors of nodal active and reactive power injections. The reference bus and PV buses are equipped with excitors or controllable reactive power compensators to maintain the pre-scheduled voltage magnitudes, and the reference bus is with fixed phase angle, typically 0. The state variables are arranged in the following sequence: $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\mathcal{R}}, \boldsymbol{\theta}_{\mathcal{S}}, \boldsymbol{\theta}_{\mathcal{L}})$ and $\boldsymbol{v} = (\boldsymbol{v}_{\mathcal{R}}, \boldsymbol{v}_{\mathcal{S}}, \boldsymbol{v}_{\mathcal{L}})$. Other variables and coefficient matrices are ordered accordingly. By partially inverting (4-3) and taking a incremental form, we have

$$\begin{bmatrix} \Delta\boldsymbol{\theta}_{\mathcal{S}} \\ \Delta\boldsymbol{\theta}_{\mathcal{L}} \\ \Delta\boldsymbol{v}_{\mathcal{L}} \end{bmatrix} = -\boldsymbol{N}^{-1} \begin{bmatrix} \Delta\boldsymbol{p}_{\mathcal{S}} \\ \Delta\boldsymbol{p}_{\mathcal{L}} \\ \Delta\boldsymbol{q}_{\mathcal{L}} \end{bmatrix} - \boldsymbol{N}^{-1}\boldsymbol{H} \begin{bmatrix} 0 \\ \Delta\boldsymbol{v}_{\mathcal{R}} \\ \Delta\boldsymbol{v}_{\mathcal{S}} \end{bmatrix} \tag{4-5}$$

$$\begin{bmatrix} \Delta\boldsymbol{q}_{\mathcal{R}} \\ \Delta\boldsymbol{q}_{\mathcal{S}} \end{bmatrix} = \boldsymbol{L}\boldsymbol{N}^{-1} \begin{bmatrix} \Delta\boldsymbol{p}_{\mathcal{S}} \\ \Delta\boldsymbol{p}_{\mathcal{L}} \\ \Delta\boldsymbol{q}_{\mathcal{L}} \end{bmatrix} + (\boldsymbol{L}\boldsymbol{N}^{-1}\boldsymbol{H} - \boldsymbol{M}) \begin{bmatrix} 0 \\ \Delta\boldsymbol{v}_{\mathcal{R}} \\ \Delta\boldsymbol{v}_{\mathcal{S}} \end{bmatrix} \tag{4-6}$$

where

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{B}'_{\mathcal{S}\mathcal{R}} & -\boldsymbol{G}_{\mathcal{S}\mathcal{R}} & -\boldsymbol{G}_{\mathcal{S}\mathcal{S}} \\ \boldsymbol{B}'_{\mathcal{L}\mathcal{R}} & -\boldsymbol{G}_{\mathcal{L}\mathcal{R}} & -\boldsymbol{G}_{\mathcal{L}\mathcal{S}} \\ \boldsymbol{G}_{\mathcal{L}\mathcal{R}} & \boldsymbol{B}_{\mathcal{L}\mathcal{R}} & \boldsymbol{B}_{\mathcal{L}\mathcal{S}} \end{bmatrix} \tag{4-7}$$

$$\boldsymbol{N} = \begin{bmatrix} \boldsymbol{B}'_{\mathcal{S}\mathcal{S}} & \boldsymbol{B}'_{\mathcal{S}\mathcal{L}} & -\boldsymbol{G}_{\mathcal{S}\mathcal{L}} \\ \boldsymbol{B}'_{\mathcal{L}\mathcal{S}} & \boldsymbol{B}'_{\mathcal{L}\mathcal{L}} & -\boldsymbol{G}_{\mathcal{L}\mathcal{L}} \\ \boldsymbol{G}_{\mathcal{L}\mathcal{S}} & \boldsymbol{G}_{\mathcal{L}\mathcal{L}} & \boldsymbol{B}_{\mathcal{L}\mathcal{L}} \end{bmatrix} \tag{4-8}$$

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{G}_{\mathcal{RR}} & \boldsymbol{B}_{\mathcal{RR}} & \boldsymbol{B}_{\mathcal{RS}} \\ \boldsymbol{G}_{\mathcal{SR}} & \boldsymbol{B}_{\mathcal{SR}} & \boldsymbol{B}_{\mathcal{SS}} \end{bmatrix} \tag{4-9}$$

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{G}_{\mathcal{RS}} & \boldsymbol{G}_{\mathcal{RL}} & \boldsymbol{B}_{\mathcal{RL}} \\ \boldsymbol{G}_{\mathcal{SS}} & \boldsymbol{G}_{\mathcal{SL}} & \boldsymbol{B}_{\mathcal{SL}} \end{bmatrix}. \tag{4-10}$$

Given a nominal operation point $\boldsymbol{x} = (\boldsymbol{\theta}, \boldsymbol{v}, \boldsymbol{p}^g, \boldsymbol{q}^g)$, equation (4-5) and (4-6) constitute the explicit formulas for calculating system response under VRE and load uncertainties. Due to the implementation of AVR, (4-5) and (4-6) are further simplified as $\Delta\boldsymbol{v}_{\mathcal{R}} = \Delta\boldsymbol{v}_{\mathcal{S}} = \boldsymbol{0}$. The method in this chapter can be used to consider any form of power injection uncertainties. But for ease of notation and without loss of generality, in the description of our method, we only consider the VRE uncertainties which are characterized by the random forecasting errors $\tilde{\boldsymbol{\zeta}}$ of VRE active power generation. The VREs are assumed to maintain a fixed power factor $\cos\phi$ at the points of connection. Under the regulation of AGC, the generating units respond affinely to the total forecasting errors. As a result, the incremental active and reactive power injections in equation (4-5) and (4-6) become random variables (with tildes) as follows:

$$\Delta\tilde{\boldsymbol{p}}_{\mathcal{S}} = -(\boldsymbol{1}^\top\tilde{\boldsymbol{\zeta}})\boldsymbol{\alpha}_{\mathcal{S}} + \tilde{\boldsymbol{\zeta}}_{\mathcal{S}} \tag{4-11a}$$

$$\Delta\tilde{\boldsymbol{p}}_{\mathcal{L}} = \tilde{\boldsymbol{\zeta}}_{\mathcal{L}} \tag{4-11b}$$

$$\Delta\tilde{\boldsymbol{q}}_{\mathcal{L}} = \sigma\boldsymbol{\zeta}_{\mathcal{L}} \tag{4-11c}$$

where $\sigma = \sin\phi/\cos\phi$ with $\cos\phi$ being the mandated power factor of wind farms.

In summary, submitting (4-11) into (4-5)(4-6) along with (4-4) and re-arranging the equations in an orgranized form, we obtain the following expression for system responses:

$$\tilde{\boldsymbol{v}}_{\mathcal{L}} = (\boldsymbol{1}^\top\tilde{\boldsymbol{\zeta}})\boldsymbol{A}^v\boldsymbol{\alpha} + \boldsymbol{B}^v\tilde{\boldsymbol{\zeta}} + \boldsymbol{v}_{\mathcal{L}} \tag{4-12a}$$

$$\tilde{\boldsymbol{q}}_{\mathcal{R}\cup\mathcal{S}} = (\boldsymbol{1}^\top\tilde{\boldsymbol{\zeta}})\boldsymbol{A}^q\boldsymbol{\alpha} + \boldsymbol{B}^q\tilde{\boldsymbol{\zeta}} + \boldsymbol{q}_{\mathcal{R}\cup\mathcal{S}} \tag{4-12b}$$

$$\tilde{\boldsymbol{f}} = (\boldsymbol{1}^\top\tilde{\boldsymbol{\zeta}})\boldsymbol{A}^f\boldsymbol{\alpha} + \boldsymbol{B}^f\tilde{\boldsymbol{\zeta}} + \boldsymbol{f} \tag{4-12c}$$

where $\boldsymbol{A}^i, \boldsymbol{B}^i, i = v, q, f$ are all constant matrices decided by network parameters and nominal line MW flow $\boldsymbol{f}$ is given by (4-4). To facilitate further discussion, we rewrite (4-12) in a component-wise form as follows:

$$\tilde{v}_i = v_i(\boldsymbol{x}, \boldsymbol{1}^\top\tilde{\boldsymbol{\zeta}}, \boldsymbol{B}^v_{i:}\tilde{\boldsymbol{\zeta}}), i \in \mathcal{L} \tag{4-13a}$$

$$\tilde{q}_j = q_i(\boldsymbol{x}, \boldsymbol{1}^\top\tilde{\boldsymbol{\zeta}}, \boldsymbol{B}^q_{j:}\tilde{\boldsymbol{\zeta}}), j \in \mathcal{R}\cup\mathcal{S} \tag{4-13b}$$

$$\tilde{f}_k = f_k(\boldsymbol{x}, \boldsymbol{1}^\top\tilde{\boldsymbol{\zeta}}, \boldsymbol{B}^f_{k:}\tilde{\boldsymbol{\zeta}}), k = 1, 2, \cdots, n_l \tag{4-13c}$$

where $\boldsymbol{B}^v_{i:}$ denotes the ith row of matrix $\boldsymbol{B}^v$ and this notation extends to other matrices as well; the explicit formulas for $v_i(\cdot)$, $q_j(\cdot)$ and $f_k(\cdot)$ can be easily seen from (4-12).

4.2.3 Problem Formulation of Distributionally Robust Chance-constrained Approximate AC-OPF

Based on the discussion in the last subsection, we formulate the distributionally robust chance-constrained approximate AC-OPF as follows:

$$\min_{\boldsymbol{x}} \sup_{\mathbb{P}\in\hat{\mathcal{P}}_N} \mathbb{E}_{\mathbb{P}} \left\{ \sum_{i=1}^{n_b} f_i \left(p_i^g - (\mathbf{1}^\top \tilde{\boldsymbol{\zeta}}) \alpha_i \right) + \overline{\boldsymbol{c}}^\top \overline{\boldsymbol{r}} + \underline{\boldsymbol{c}}^\top \underline{\boldsymbol{r}} \right\} \tag{4-14}$$

subject to

$$\boldsymbol{p}^g + \boldsymbol{p}^w - \boldsymbol{p}^l = \mathscr{P}(\boldsymbol{\theta}, \boldsymbol{v}) \tag{4-15a}$$

$$\boldsymbol{q}^g + \boldsymbol{q}^w - \boldsymbol{q}^l = \mathscr{Q}(\boldsymbol{\theta}, \boldsymbol{v}) \tag{4-15b}$$

$$\underline{\boldsymbol{v}} \le \boldsymbol{v} \le \overline{\boldsymbol{v}} \tag{4-15c}$$

$$\mathbf{1}^\top \boldsymbol{\alpha} = 1,\ \boldsymbol{\alpha} \ge \mathbf{0} \tag{4-15d}$$

$$\boldsymbol{p}^g - \underline{\boldsymbol{r}} \ge \underline{\boldsymbol{p}}^g,\ \underline{\boldsymbol{r}} \ge \mathbf{0} \tag{4-15e}$$

$$\boldsymbol{p}^g + \overline{\boldsymbol{r}} \le \overline{\boldsymbol{p}}^g,\ \overline{\boldsymbol{r}} \ge \mathbf{0} \tag{4-15f}$$

$$\inf_{\mathbb{P}\in\hat{\mathcal{P}}_N} \mathbb{P}\left\{ -\underline{\boldsymbol{r}} \le -(\mathbf{1}^\top \tilde{\boldsymbol{\zeta}}) \boldsymbol{\alpha} \le \overline{\boldsymbol{r}} \right\} \ge 1 - \rho_1 \tag{4-15g}$$

$$\forall i \in \mathcal{L},\ j \in \mathcal{R} \cup \mathcal{S},\ k = 1, 2, \cdots, n_l:$$

$$\inf_{\mathbb{P}\in\hat{\mathcal{P}}_N} \mathbb{P}\left\{ \underline{v}_i \le \tilde{v}_i \le \overline{v}_i \right\} \ge 1 - \rho_2 \tag{4-15h}$$

$$\inf_{\mathbb{P}\in\hat{\mathcal{P}}_N} \mathbb{P}\left\{ \underline{q}_j \le \tilde{q}_j \le \overline{q}_j \right\} \ge 1 - \rho_3 \tag{4-15i}$$

$$\inf_{\mathbb{P}\in\hat{\mathcal{P}}_N} \mathbb{P}\left\{ \underline{f}_k \le \tilde{f}_k \le \overline{f}_k \right\} \ge 1 - \rho_4 \tag{4-15j}$$

where $\boldsymbol{x} = (\boldsymbol{v}, \boldsymbol{\theta}, \boldsymbol{p}^g, \boldsymbol{q}^g, \boldsymbol{\alpha}, \overline{\boldsymbol{r}}, \underline{\boldsymbol{r}})$; $\mathcal{P}_N$ is an ambiguity set of probability distributions constructed from historical data and its explicit expression will be discussed in the next section. The objective function (4-14) is the worst-case expectation of generation and reserve costs in which the cost function $f_i(\cdot), i = 1, \cdots, n_g$ are convex quadratic functions of the form:

$$f_i(x) = c_{i2} x^2 + c_{i1} x + c_{i0}. \tag{4-16}$$

with $c_{i2}, c_{i1}, c_{i0} \ge 0$. Nolinear equality constraint (4-15a) and (4-15b) ensure the active and reactive power balance at the nominal opeartion point where the VRE generation exactly match the forecasting values. Constraint (4-15c) sets the voltage magnitude limits for all buses at the nominal opeartion point and also for the reference and PV buses under perturbed conditions. Constraint (4-15d) enforces the basic requirement for participation factors in the AGC system. Constraint (4-15e) and (4-15f) guarantee the reserve availability considering the output limits of generating units. The adequacy of downward and upward regulating reverse is ensured by distributionally robust chance constraint (4-15g). Similarly, constraint (4-15h) ensures the

voltage quality for PQ buses, and constraint (4-15i) safeguards the adequancy of reactive power for AVR at generator buses. Finally, constraint (4-15j) guarantees the adequacy of transmission line capacity.

Remark 4.1： The OPF formulation (4-14)(4-15) combines the exact nonlinear power flow equations (4-1) and the linear approximate system response model (4-13). On the one hand, this new problem formulation inherits the basic feature of conventional AC-OPF by using the exact nonlinear power flow to govern the system state at the nominal condition. On the other hand, the linear approximate power flow response model is employed to describe the system behavior when VRE generation deviates from forecasting values, which makes it possible to develop efficient techniques to deal with the stochastic formulation (4-14)(4-15g)∼(4-15j). Of course, though more tractable, the linear power flow response model still brings inaccuracies in evaluating the operational costs, voltages, reactive power and line flows. The accuracy of this approach will be assessed by comparative numerical studies in section 4.5. □

Remark 4.2： Note that the dimension of random variable $\tilde{\boldsymbol{\zeta}}$ is equal to the number of VREs installed across the network. It could be prohibitively high for numerical computation in a real-world large-scale power system with distributed wind farms and PVs. Fortunately, the simple affine policy employed in the AGC scheme enables significant reduction in the dimension of the random variable considered in model (4-14)(4-15). Specifically, in the objective function (4-14) and constraint (4-15g), we only need to consider the 1-dimensional random variable $\mathbf{1}^\top\tilde{\boldsymbol{\zeta}}$. For the constraint (4-15h), it is only required to consider the 2-dimensional random variable $(\mathbf{1}^\top\tilde{\boldsymbol{\zeta}}, \boldsymbol{B}^v_{i:}\tilde{\boldsymbol{\zeta}})$ in light of equation (4-13). Similarly, the 2-dimensional random variable $(\mathbf{1}^\top\tilde{\boldsymbol{\zeta}}, \boldsymbol{B}^f_{i:}\tilde{\boldsymbol{\zeta}})$ needs to be taken into account in constraint (4-15j). When the historical data of $\tilde{\boldsymbol{\zeta}}$ is available at hand, it is equivalent to have the historical data for the 1- or 2-dimensional random variables described above. In summary, no matter how many uncertain renewable sources are installed in the system, we only need to deal with 1- or 2-dimensional random variables which significantly improve the scalability of the proposed method. □

Remark 4.3： In a conventional stochastic chance-constrained programming, the probability distribution $\mathbb{P}$ is assumed to be known apriori. However, in practice, the probability distribution of $\tilde{\xi}$ is unknown and only some historical data is available. Theoretically, the precise knowledge of the probability distribution cannot be obtained from finite data. Therefore, the probability distribution is ambiguous. Nonetheless, the historical data does provide us some reliable probabilistic information based on which we can construct an ambiguity set $\mathcal{P}_N$, i.e. a set of probability distributions consistent with the observed historical data. Therefore, the objective function (4-14) and the chance-constraints (4-15g)∼(4-15j) consider the worst-case distribution in the ambiguity set $\mathcal{P}_N$ to ensure the strategy performs well under the ambiguity of distribution. □

4.3 Ambiguity Set Using Wasserstein Metric

4.3.1 Wasserstein Metric and Ambiguity Set

As shown in equation (4-12), the system responses are the functions of some random variable $\tilde{\boldsymbol{\xi}}$. Whether evaluating the expected generation costs or assessing the reliability of system operation, we need to have a knowledge about the probability distribution $\mathbb{P}$ of random variable $\tilde{\boldsymbol{\xi}}$ under consideration. Unfortunately, in the real-life application, we only have a finite set of historical data and the precise characterization of $\mathbb{P}$ can never be extracted from finite samples. In other words, the underlying true distribution $\mathbb{P}$ is ambiguous. However, according to the historical sample set $\{\hat{\boldsymbol{\xi}}^{(1)}, \hat{\boldsymbol{\xi}}^{(2)}, \cdots, \hat{\boldsymbol{\xi}}^{(N)}\}$ at hand, we can construct an empirical distribution $\hat{\mathbb{P}}_N = \frac{1}{N}\sum_{k=1}^{N} \delta_{\hat{\boldsymbol{\xi}}^{(k)}}$ which acts as an estimation of the true distribution $\mathbb{P}$. Intuitively, $\hat{\mathbb{P}}_N$ converges to $\mathbb{P}$ as $N \to \infty$, i.e. the "distance" between $\hat{\mathbb{P}}_N$ and $\mathbb{P}$ becomes smaller when more data is available. One of the "distances" to establish the convergence of $\hat{\mathbb{P}}_N$ to $\mathbb{P}$ is the Wasserstein metric defined as follows:

Definition 4.1 (Wasserstein Metric)： For any probability distribution $\mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{P}(\Xi)$, the Wasserstein metric can be defined through

$$W(\mathbb{Q}_1,\mathbb{Q}_2) = \inf_{\Pi}\left\{ \int_{\Xi^2} \|\xi_1 - \xi_2\| \Pi(d\xi_2, d\xi_2) : \begin{array}{l} \Pi \text{ is a joint distribution of } \xi_1 \text{ and } \xi_2 \\ \text{with marginals } \mathbb{Q}_1 \text{ and } \mathbb{Q}_2 \end{array} \right\}. \tag{4-17}$$

$\Diamond$

where $\mathcal{P}(\Xi)$ denotes the set of all probability distributions with support Ξ; $\|\cdot\|$ can be any norm in $\mathbb{R}^n$, and we use l_1 norm $\|\cdot\|_1$ in this chapter for its superior numerical tractability in DRO. Hence, we have $W(\hat{\mathbb{P}}_N, \mathbb{P}) \leq \epsilon(N)$ where $\epsilon(\cdot)$ is some sample-dependent monotone function decreasing to 0 as N tends to infinity, and the explicit formula for $\epsilon(\cdot)$ will be discussed in the next subsection. Therefore, given a historical data set with N samples, we know that the true distribution $\mathbb{P}$ belongs to the following set:

$$\hat{\mathcal{P}}_N = \left\{ \mathbb{P} \in \mathcal{P}(\Xi) : W(\mathbb{P}, \hat{\mathbb{P}}_N) \leq \epsilon(N) \right\} \tag{4-18}$$

which is called a *ambiguity set* of the underlying true distribution. Note that $\hat{\mathcal{P}}_N$ is the Wasserstein ball of radius $\epsilon(N)$ centered at the empirical distribution $\hat{\mathbb{P}}_N$. It represents the reliable information about the true distribution $\mathbb{P}$ observed from the N historical samples at hand.

4.3.2 Selection of Wasserstein Radius $\epsilon(N)$

The radius of the Wasserstein ball in the ambiguity set (4-18) is crucial for the performance of the distributionally robust chance-constrained optimization problem (4-14)(4-15).

Following the discussion in last subsection, $\epsilon(N)$ is a decreasing function, as small as possible, to bound $W(\hat{\mathbb{P}}_N, \mathbb{P})$ from above. One possible choice for $\epsilon(N)$ was given in [113] as follows:

$$\epsilon(N) = D\sqrt{\frac{2}{N}\log\left(\frac{1}{1-\beta}\right)}, \tag{4-19}$$

where β is a confidence level and D is the diameter of the support of the random variable. However, in our numerical experience, we find this choice of $\epsilon(N)$ overly conservative for the DRO to gain noticable advantage over the conventional RO with reasonable amount of data.

To improve the formula (4-19), let's review the derivation of (4-19) presented in [113, 114]. In the proof of proposition 3 in [114], it is shown that for any $\delta > 0$

$$\mathbb{P}\{W(\hat{\mathbb{P}}_N, \mathbb{P}) \geq \epsilon\} \leq \exp\left(-N \inf_{\mathbb{Q}: W(\mathbb{Q},\mathbb{P}) \geq \epsilon} (H(\mathbb{Q}|\mathbb{P}) - \delta)\right) \tag{4-20}$$

where $H(\mathbb{Q}|\mathbb{P})$ is called the Kullback information of $\mathbb{Q}$ w.r.t. $\mathbb{P}$ (see [115] for more detail). Corollary 2.4 in [115] shows

$$H(\mathbb{Q}|\mathbb{P}) \geq W(\mathbb{Q}, \mathbb{P})^2 / C^2, \ \forall \mathbb{Q} \in \mathcal{P}(\Xi) \tag{4-21}$$

where

$$C = 2 \inf_{\boldsymbol{\xi}_0 \in \Xi, \alpha > 0} \left(\frac{1}{2\alpha}(1 + \ln \mathbb{E}_{\mathbb{P}}[e^{\alpha \|\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_0\|_1^2}])\right)^{1/2} \tag{4-22}$$

with a particular case $C = \sqrt{2}D$ (particular case 2.5 in [115]). Plugging (4-21) into RHS of (4-20) and taking $\delta \to 0$, we have

$$\mathbb{P}\{W(\hat{\mathbb{P}}_N, \mathbb{P}) \leq \epsilon\} \geq 1 - \exp\left(-N\frac{\epsilon^2}{C^2}\right). \tag{4-23}$$

Equating the RHS of (4-23) to β leads to

$$\epsilon(N) = C\sqrt{\frac{1}{N}\log\left(\frac{1}{1-\beta}\right)}. \tag{4-24}$$

Note that if the special case $C = \sqrt{2}D$ is adopted, (4-24) turns into (4-19). However, $\sqrt{2}D$ is much larger than (4-22) in practice, which is one of the major sources of conservatism for (4-19). Instead of adopting the special case $C = \sqrt{2}D$, we directly use (4-22) by estimating its value from data. From (4-22), we have

$$\begin{aligned} C \leq & 2 \inf_{\alpha > 0} \left(\frac{1}{2\alpha}(1 + \ln \mathbb{E}_{\mathbb{P}}[e^{\alpha \|\tilde{\boldsymbol{\xi}} - \hat{\boldsymbol{\mu}}\|_1^2}])\right)^{1/2} \\ \approx & 2 \inf_{\alpha > 0} \left(\frac{1}{2\alpha}\left(1 + \ln\left(\frac{1}{N}\sum_{k=1}^{N} e^{\alpha \|\hat{\boldsymbol{\xi}}^{(k)} - \hat{\boldsymbol{\mu}}\|_1^2}\right)\right)\right)^{1/2} \end{aligned} \tag{4-25}$$

where $\hat{\boldsymbol{\mu}}$ denotes the sample mean and the minimization over α can be easily done by the bisection search method.

4.4 Solution Approach

To solve the distributionally robust voltage-concerned chance-constrained AC-OPF (4-14)(4-15), we need to evaluate the worst-case expected costs in (4-14) and reformulate the distributionally robust chance constraints (4-15g)∼(4-15j). The underlying engine for the reformulation is the strong duality result developed in [110]. We tailor the result to our problem setting as the following lemma.

Lemma 4.1 ([110])： Given a random variable $\tilde{\boldsymbol{\xi}} \in \mathbf{R}^m$ with closed and convex support Ξ, the Wasserstein ball $\mathcal{B}_\epsilon(\hat{\mathbb{P}}_\xi)$ is constructed from sample set $\{\hat{\boldsymbol{\xi}}_1, \hat{\boldsymbol{\xi}}_2, \cdots, \hat{\boldsymbol{\xi}}_N\}$. If the loss function $l(\tilde{\boldsymbol{\xi}})$ is upper semi-continuous, the worst-case expectation

$$\begin{aligned} &\sup_{\mathbb{P}\in\hat{\mathcal{P}}_N} \mathbb{E}_{\mathbb{P}}\left\{l(\tilde{\boldsymbol{\xi}})\right\} \\ =&\inf_{\lambda\geq 0}\left\{\lambda\cdot\epsilon + \frac{1}{N}\sum_{k=1}^{N}\sup_{\boldsymbol{\xi}\in\Xi}\left(l(\boldsymbol{\xi}) - \lambda\|\boldsymbol{\xi}-\hat{\boldsymbol{\xi}}_k\|_1\right)\right\}. \end{aligned} \tag{4-26}$$

□

In our implementation, we estimate the support Ξ from the sample set as

$$\Xi = \{\boldsymbol{\xi}\in\mathbf{R}^m | -\sigma_{\max}\mathbf{1} \leq \hat{\boldsymbol{\Sigma}}^{-1/2}(\boldsymbol{\xi}-\hat{\boldsymbol{\mu}}) \leq \sigma_{\max}\mathbf{1}\} \tag{4-27}$$

where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are the sample mean and sample covariance, and $\sigma_{\max} = 10$ in our practice.

4.4.1 Evaluation of Worst-case Costs

By noticing (4-16), the cost function inside the $\mathbb{E}_{\mathbb{P}}\{\cdot\}$ operator in (4-14) can be written as follows:

$$\eta(\boldsymbol{x}, \tilde{\omega}) = c_2\tilde{\omega}^2 - c_1\tilde{\omega} + c_0 \tag{4-28}$$

where

$$\begin{cases} c_2 = \sum_{i=1}^{n_g} c_{i2}\alpha_i^2 \\ c_1 = \sum_{i=1}^{n_g} 2c_{i2}p_i\alpha_i + c_{i1}\alpha_i \\ c_0 = \sum_{i=1}^{n_g}(c_{i2}p_i^2 + c_{i1}p_i + c_{i0}) + \overline{\boldsymbol{c}}^\top\overline{\boldsymbol{r}} + \underline{\boldsymbol{c}}\top\underline{\boldsymbol{r}}. \end{cases} \tag{4-29}$$

with $\tilde{\omega} = \mathbf{1}^\top\tilde{\boldsymbol{\zeta}}$. Since $c_{i2} \geq 0$, we have $c_2 \geq 0$. It follows that $\eta(\boldsymbol{x}, \cdot)$ is a convex quadratic function. In addition, $\eta(\cdot, \tilde{\omega})$ is also a convex quadratic function because it is the sum of convex quadratic funtions $\{f_i\}_{i=1}^{n_g}$.

Given sample set $\{\hat{\omega}_1, \hat{\omega}_2, \cdots, \hat{\omega}_N\}$ of random variable $\tilde{\omega}$ and its support $[\underline{\omega}, \overline{\omega}]$, we can evaluate the worst-case costs in (4-14) using Lemma 4.1 as follows:

$$\sup_{\mathbb{P}\in\hat{\mathcal{P}}_N} \mathbb{E}_{\mathbb{P}}\left\{\eta(\boldsymbol{x}, \tilde{\omega})\right\} \tag{4-30a}$$

$$
= \begin{cases} \inf\limits_{\lambda \geq 0, s \in \mathbf{R}^N} \lambda \cdot \epsilon + \frac{1}{N} \sum_{k=1}^{N} s_k \\ \text{s.t. } \sup\limits_{\underline{\omega} \leq \omega \leq \overline{\omega}} (\eta(\boldsymbol{x}, \omega) - \lambda |\omega - \hat{\omega}_k|) \leq s_k, \forall k \leq N \end{cases} \tag{4-30b}
$$

$$
= \begin{cases} \inf\limits_{\lambda \geq 0, s \in \mathbf{R}^N} \lambda \cdot \epsilon + \frac{1}{N} \sum_{k=1}^{N} s_k \\ \text{s.t. } \eta(\boldsymbol{x}, \underline{\omega}) + \lambda(\underline{\omega} - \hat{\omega}_k) \leq s_k, \forall k \leq N \\ \quad \eta(\boldsymbol{x}, \overline{\omega}) - \lambda(\overline{\omega} - \hat{\omega}_k) \leq s_k, \forall k \leq N \\ \quad \eta(\boldsymbol{x}, \hat{\omega}_k) \leq s_k, \forall k \leq N \end{cases} \tag{4-30c}
$$

where (4-30b) follows from a direct application of Lemma 4.1 and the introduction of epigraphical auxiliary variables $s_k, k \leq N$; equality (4-30c) comes from the observation that $\eta(\boldsymbol{x}, \omega) - \lambda|\omega - \hat{\omega}_k|$ is convex in ω in the intervals $[\underline{\omega}, \hat{\omega}_k]$ and $[\hat{\omega}_k, \overline{\omega}]$, repectively. Hence, the supremum in (4-30b) can only be attained at the boundary of the intervals, i.e. $\underline{\omega}$, $\hat{\omega}_k$ or $\overline{\omega}$. Note that the worst-case expected costs have already been transformed into a tractable deterministic formulation (4-30c) with linear objective function and convex quadratic constraints. However, problem (4-30c) has a huge computational disadvantage: the numbers of auxiliary variables and quadratic constraints are propositional to the size of historical sample set. When a large data set is at hand, the computational burden could prevent us fully exploiting the historical data.

To overcome this drawback, we replace (4-30c) with a close upper approximation which is more scalable w.r.t. the size of sample set. Let $\eta'(\boldsymbol{x}, \tilde{\omega}) = 2c_2\tilde{\omega} - c_1$ be the derivative of $\eta(\boldsymbol{x}, \tilde{\omega})$ w.r.t. $\tilde{\omega}$, and take $\lambda = \max\{\eta'(\boldsymbol{x}, \overline{\omega}), -\eta'(\boldsymbol{x}, \underline{\omega})\}$ which is the smallest value of λ in (4-30c) such that

$$
\begin{cases} \eta(\boldsymbol{x}, \underline{\omega}) + \lambda(\underline{\omega} - \omega) \leq \eta(\boldsymbol{x}, \omega), \ \forall \omega \in [\underline{\omega}, \overline{\omega}] \\ \eta(\boldsymbol{x}, \overline{\omega}) - \lambda(\overline{\omega} - \omega) \leq \eta(\boldsymbol{x}, \omega), \ \forall \omega \in [\underline{\omega}, \overline{\omega}]. \end{cases} \tag{4-31}
$$

Thus,

$$
(4-30c) \leq \begin{cases} \inf\limits_{\lambda \in \mathbf{R}} \lambda \cdot \epsilon + \frac{1}{N} \sum_{k=1}^{N} \eta(\boldsymbol{x}, \hat{\omega}_k) \\ \text{s.t. } \eta'(\boldsymbol{x}, \overline{\omega}) \leq \lambda \\ \quad -\eta'(\boldsymbol{x}, \underline{\omega}) \leq \lambda. \end{cases} \tag{4-32}
$$

Note that the numbers of decision variables and constraints in (4-32) stay unchanged when using larger sample set, therefore (4-32) is more computationally favorable than (4-30c). In addition, the optimums of (4-30c) and (4-32) have very small difference in practice and such difference diminishes as the sample size N increases.

4.4.2 Reformulation of Distributionally Robust Chance Constraints

The chance constraints (4-15g)∼(4-15j) can be written in the general form:

$$\inf_{\mathbb{P}\in\hat{\mathcal{P}}_N} \mathbb{P}\{\boldsymbol{g}(\boldsymbol{x},\tilde{\boldsymbol{\xi}}) \leq \mathbf{0}\} \geq 1-\rho. \tag{4-33}$$

where $\boldsymbol{g}$ depends linearly on decision variable $\boldsymbol{x}$ and random variable $\tilde{\boldsymbol{\xi}}$, respectively. The above constraint is in fact non-convex, so it is generally difficult to derive a tractable equivalent reformulation. In this chapter, we develop a convex conservative approximation to (4-33) through the following strategy. We seek a deterministic uncertainty set $\mathcal{U}$ such that the robust constraint

$$\boldsymbol{g}(\boldsymbol{x},\boldsymbol{\xi}) \leq \mathbf{0},\ \forall \boldsymbol{\xi} \in \mathcal{U} \tag{4-34}$$

implies the distributionally robust chance constraint (4-33). In addition, the uncertainty set $\mathcal{U}$ should possess the following features: 1) obtaining $\mathcal{U}$ is computationally cheap; 2) robust constraint (4-34) is numerically tractable and efficient; 3) uncertainty set $\mathcal{U}$ is made as small as possible.

Given a sample set $\{\hat{\boldsymbol{\xi}}^{(1)}, \hat{\boldsymbol{\xi}}^{(2)}, \cdots, \hat{\boldsymbol{\xi}}^{(N)}\}$ of random variable $\tilde{\boldsymbol{\xi}} \in \mathbf{R}^m$ with the unknown underlying true distribution $\mathbb{P}$, we can compute the sample mean $\hat{\boldsymbol{\mu}}$ and the sample covariance $\hat{\boldsymbol{\Sigma}}$. Instead of directly working with random variable $\tilde{\boldsymbol{\xi}}$, we consider its standardized version $\tilde{\boldsymbol{\vartheta}} = \hat{\boldsymbol{\Sigma}}^{-1/2}(\tilde{\boldsymbol{\xi}} - \hat{\boldsymbol{\mu}})$ with sample set $\{\hat{\boldsymbol{\vartheta}}^{(k)} = \hat{\boldsymbol{\Sigma}}^{-1/2}(\hat{\boldsymbol{\xi}}^{(k)} - \hat{\boldsymbol{\mu}})\}_{k=1,2,\cdots,N}$, illustrated in Fig. 4-3. Obviously, random variable $\tilde{\boldsymbol{\vartheta}}$ has sample mean $\mathbf{0}$, sample covariance $\mathbf{I}$ and support $\Theta = \{-\sigma_{\max}\mathbf{1} \leq \boldsymbol{\vartheta} \leq \sigma_{\max}\mathbf{1}\}$. Let $\mathbb{Q}$ and $\mathbb{Q}_N$ denote the true distribution and empirical distribution of $\tilde{\boldsymbol{\vartheta}}$, respectively. Accordingly, the ambiguity set $\hat{\mathcal{Q}}_N$ can be constructed as in (4-18). We seek a set $\mathcal{V} \subseteq \mathbf{R}^m$ such that

$$\sup_{\mathbb{Q}\in\hat{\mathcal{Q}}_N} \mathbb{Q}\{\tilde{\boldsymbol{\vartheta}} \notin \mathcal{V}\} \leq \rho. \tag{4-35}$$

Then $\mathcal{U} = \hat{\boldsymbol{\Sigma}}^{1/2}\mathcal{V} + \hat{\boldsymbol{\mu}}$ would be a desired uncertainty set needed in (4-34). To develop an efficient method to find out such $\mathcal{V}$, considering the sample uncorrelatedness and equal variance among different components of $\tilde{\boldsymbol{\vartheta}}$, we restrict the structure of $\mathcal{V}$ as the hypercube

$$\mathcal{V}(\sigma) = \{\boldsymbol{\vartheta} \in \mathbf{R}^m | -\sigma\mathbf{1} < \boldsymbol{\vartheta} < \sigma\mathbf{1}\} \tag{4-36}$$

whose side length σ needs to be minimized in order to reduce conservatism. This leads to the following optimization problem:

$$\min_{0\leq\sigma\leq\sigma_{\max}} \sigma \ \text{ s.t. } \sup_{\mathbb{Q}\in\hat{\mathcal{Q}}_N} \mathbb{Q}\{\tilde{\boldsymbol{\vartheta}} \notin \mathcal{V}(\sigma)\} \leq \rho. \tag{4-37}$$

By leveraging the worst-case probability formulation in [109] and the duality result in Lemma 4.1, problem (4-37) has a very simple deterministic reformulation which is a consequence of the following lemma.

Lemma 4.2：

$$\begin{aligned}&\sup_{\mathbb{Q}\in\hat{\mathcal{Q}}_N}\mathbb{Q}\{\tilde{\boldsymbol{\vartheta}}\notin\mathcal{V}(\sigma)\}\\=&\inf_{\lambda\geq 0}\left\{\lambda\cdot\epsilon+\frac{1}{N}\sum_{k=1}^{N}\left(1-\lambda(\sigma-\|\hat{\boldsymbol{\vartheta}}^{(k)}\|_\infty)^+\right)^+\right\}.\end{aligned}\tag{4-38}$$

□

where $(x)^+=\max(x,0)$.

Proof： Using the idea from section 5.2 of [109], for every $1\leq i\leq m$, we define

$$\underline{l}_i(\tilde{\boldsymbol{\vartheta}})=\begin{cases}1 & \text{if } \tilde{\vartheta}_i\leq-\sigma\\ -\infty & \text{otherwise}\end{cases}\tag{4-39}$$

$$\bar{l}_i(\tilde{\boldsymbol{\vartheta}})=\begin{cases}1 & \text{if } \tilde{\vartheta}_i\geq\sigma\\ -\infty & \text{otherwise .}\end{cases}\tag{4-40}$$

Further define

$$l(\tilde{\boldsymbol{\vartheta}})=\max\left\{\max_{1\leq i\leq m}\underline{l}_i(\tilde{\boldsymbol{\vartheta}}),\max_{1\leq i\leq m}\bar{l}_i(\tilde{\boldsymbol{\vartheta}}),0\right\}.\tag{4-41}$$

It is easy to see that $l(\tilde{\boldsymbol{\vartheta}})$ is the indicator function of the compliment of $\mathcal{V}(\sigma)$, so $\mathbb{Q}\{\tilde{\boldsymbol{\vartheta}}\notin\mathcal{V}(\sigma)\}=\mathbb{E}_{\mathbb{Q}}\{l(\tilde{\boldsymbol{\vartheta}})\}$. In addition, $l(\tilde{\boldsymbol{\vartheta}})$ is the maximum of upper-semicontinuous functions, which implies $l(\tilde{\boldsymbol{\vartheta}})$ itself is upper-semicontinuous. Therefore, Lemma 4.1 applies:

$$\sup_{\mathbb{Q}\in\hat{\mathcal{Q}}_N}\mathbb{Q}\{\tilde{\boldsymbol{\vartheta}}\notin\mathcal{V}(\sigma)\}=\sup_{\mathbb{Q}\in\hat{\mathcal{Q}}_N}\mathbb{E}_{\mathbb{Q}}\{l(\tilde{\boldsymbol{\vartheta}})\}\tag{4-42a}$$

$$=\begin{cases}\inf\limits_{\lambda\geq 0,\boldsymbol{s}\in\mathbb{R}^N}\ \lambda\cdot\epsilon+\frac{1}{N}\sum_{k=1}^{N}s_k\\ \text{s.t. }\forall 1\leq k\leq N:\\ \quad\sup\limits_{\boldsymbol{\vartheta}\in\Theta}\left(l(\boldsymbol{\vartheta})-\lambda\|\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}^{(k)}\|_1\right)\leq s_k\end{cases}\tag{4-42b}$$

$$=\begin{cases}\inf\limits_{\lambda\geq 0,\boldsymbol{s}\geq\mathbf{0}}\ \lambda\cdot\epsilon+\frac{1}{N}\sum_{k=1}^{N}s_k\\ \text{s.t. }\forall 1\leq k\leq N,1\leq i\leq m:\\ \quad\sup\limits_{\boldsymbol{\vartheta}\in\Theta}\left(\underline{l}_i(\tilde{\boldsymbol{\vartheta}})-\lambda\|\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}^{(k)}\|_1\right)\leq s_k\\ \quad\sup\limits_{\boldsymbol{\vartheta}\in\Theta}\left(\bar{l}_i(\tilde{\boldsymbol{\vartheta}})-\lambda\|\boldsymbol{\vartheta}-\hat{\boldsymbol{\vartheta}}^{(k)}\|_1\right)\leq s_k\end{cases}\tag{4-42c}$$

$$=\begin{cases}\inf\limits_{\lambda\geq 0,\boldsymbol{s}\geq\mathbf{0}}\ \lambda\cdot\epsilon+\frac{1}{N}\sum_{k=1}^{N}s_k\\ \text{s.t. }\forall 1\leq k\leq N,1\leq i\leq m:\\ \quad 1-\lambda(\hat{\vartheta}_i^{(k)}+\sigma)^+\leq s_k\\ \quad 1-\lambda(\sigma-\hat{\vartheta}_i^{(k)})^+\leq s_k\end{cases}\tag{4-42d}$$

$$= \inf_{\lambda \geq 0} \left\{ \lambda \cdot \epsilon + \frac{1}{N} \sum_{k=1}^{N} \left(1 - \lambda(\sigma - \|\hat{\boldsymbol{\vartheta}}^{(k)}\|_\infty)^+\right)^+ \right\} \tag{4-42e}$$

where (4-42b) is a direct application of Lemma 4.1; (4-42c) uses the the definition of $l(\tilde{\boldsymbol{\vartheta}})$ given in (4-41); (4-42d) explicitly evaluates the supremums in (4-42c) by exploiting the definitions of Θ, (4-39)(4-40) and the l_1 norm; some direct calculation of (4-42d) leads to (4-42e) which completes the proof. ■

Using Lemma 4.2, problem (4-37) is equivalent to

$$\min_{0 \leq \lambda, 0 \leq \sigma \leq \sigma_{\max}} \sigma \ \text{ s.t. } \ h(\sigma, \lambda) \leq \rho. \tag{4-43}$$

where $h(\sigma, \lambda) = \lambda\epsilon + \frac{1}{N} \sum_{k=1}^{N} \left(1 - \lambda(\sigma - \|\hat{\boldsymbol{\vartheta}}^{(k)}\|_\infty)^+\right)^+$.

Note that the worst-case probability (4-38) is non-decreasing in σ, therefore problem (4-37), equivalently problem (4-43), has a unique minimum. Although problem (4-43) is non-smooth, it only involves two 1-dimensional decision variables. We can design a nested bisection search method to quickly locate the optimal solution. The method is summarized in Algorithm 1 in which the function `bisearch`$(f(\cdot), a, b)$ returns the minimum of $f(\cdot)$ in the interval $[a, b]$ by performing a bisection search. Further note that $h(\sigma, \lambda)$ is convex in λ for fixed σ, so the bisection search in step 4 of Algorithm 1 is well-defined. Since Algrithm 1 only involves function evaluations, it solves problem (4-43) very efficiently.

Algorithm 1 Nested Bisection Search

1: Initialize $\underline{\sigma} = 0$, $\overline{\sigma} = \sigma_{\max}$;
2: **while** ($\overline{\sigma} - \underline{\sigma} > 10^{-4}$) **do**
3: $\quad \sigma = (\overline{\sigma} + \underline{\sigma})/2$;
4: $\quad \gamma =$`bisearch`$(h(\sigma, \cdot)$,0,100);
5: $\quad$ **if** $\gamma > \rho$ **then**
6: $\quad\quad \underline{\sigma} = \sigma$;
7: $\quad$ **else**
8: $\quad\quad \overline{\sigma} = \sigma$;
9: $\quad$ **end if**
10: **end while**
11: Output $\sigma = \overline{\sigma}$.

After determining the optimal σ, the hypercube $\mathcal{V}(\sigma)$ can be expressed as the convex hull of its vertices. For 1-dimensional random variable, $\mathcal{V}(\sigma) = \text{conv}(\{-\sigma, \sigma\})$, and for 2-dimensional random variable, $\mathcal{V}(\sigma) = \text{conv}(\{(\pm\sigma, \pm\sigma)\})$. In general, $\mathcal{V}(\sigma) = \text{conv}(\{\boldsymbol{v}^{(1)}, \cdots, \boldsymbol{v}^{(2^m)}\})$. Hence the desired uncertainty set $\mathcal{U}$ takes the form

$$\mathcal{U} = \text{conv}(\{\boldsymbol{u}^{(1)}, \cdots, \boldsymbol{u}^{(2^m)}\}) \tag{4-44}$$

where $\boldsymbol{u}^{(i)} = \hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{v}^{(i)} + \hat{\boldsymbol{\mu}}$, $1 \leq i \leq 2^m$. Thus, noticing the linear dependence of $\boldsymbol{g}$ on $\boldsymbol{\xi}$, the robust constraint (4-34) is equivalent to

$$\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{u}^{(i)}) \leq \boldsymbol{0}, \ 1 \leq i \leq 2^m \tag{4-45}$$

which is a set of linear inequality constraints on decision variable $\boldsymbol{x}$. As a result, the distributionally robust chance constraint (4-33) can be safely approximated by the linear constraints (4-45). Note that we only need to deal with the cases of $m = 1$ and $m = 2$ as commented in Remark 2, therefore constraint (4-45) does not bring any scalability problems and can be handled very efficiently. The whole procedure is illustrated in Fig. 4-3.

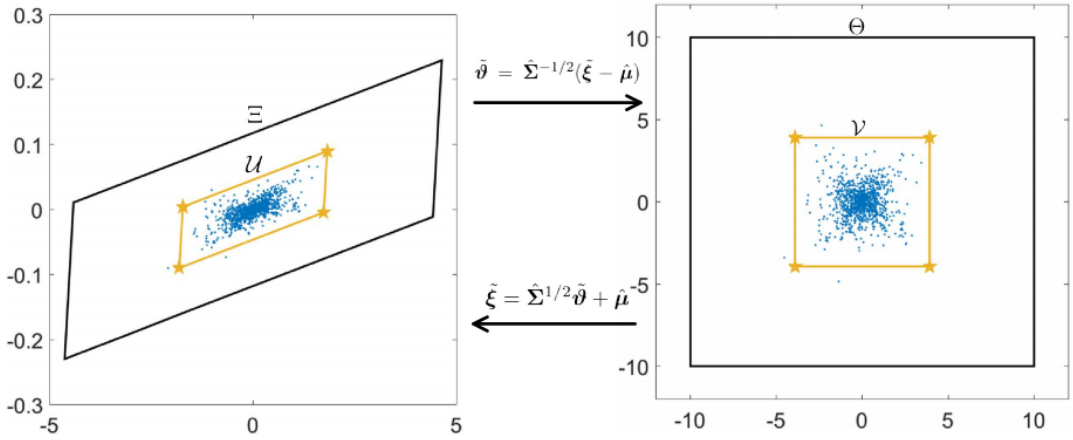


Figure 4-3 The procedure to determine the uncertainty set $\mathcal{U}$ in two dimension.

To sum up, replacing the worst-case expected costs in (4-14) with (4-32), taking the place of chance constraints (4-15g)∼(4-15j) with their corresponding deterministic constraint (4-45), we obtain a deterministic optimization problem with quadratic objective function (4-15), nonlinear equality constraints (4-15a)(4-15b), and linear constraints (4-15c)∼(4-15j). In fact, this problem structure can be seen as an extension of conventional AC-OPF by adding decision varibles $(\boldsymbol{\alpha}, \overline{\boldsymbol{r}}, \underline{\boldsymbol{r}})$ and a set of linear constraints. Therefore, the problem can be solved (at least to local optimality) by a mature interior point method (IPM) solver.

4.5 Numerical Results

This section reports numerical results. We will first discuss some implementation issues of the proposed method and the related benchmarking approaches. Then the case studies on IEEE 14 and 118 bus systems will be presented. The tests on 14-bus system emphasize on validity and necessity of the proposed approximate AC-OPF formulation. Whereas the studies on 118-bus system focus on the features of the DRO approach with the comparison to other methods to deal with uncertainties.

4.5.1 Implementation and Benchmarking

The MATPOWER [116] offers a very convenient way to implement our problem formulation. It employs an extensible OPF structure [117] to allow the user to modify and augment the problem formulation without rewriting the portions that are shared with the conventional AC-OPF. Our problem formulation fits well into this structure and the modification needed can be easily done with the help of YALMIP [118]. We choose KNITRO [119] as the default IPM solver. Further note that voltage magnitude (4-15h), reactive power (4-15i) and line flow (4-15j) constraints are only active at a fraction of buses, generators and lines in practical systems. To improve the computational efficiency, we employ the successive constraint enforcement scheme [75] which relaxes all the constraints (4-15h)∼(4-15j) initially and subsequently adds back those that are violated at the solution of the relaxed model.

In order to comprehensively assess the performance of the proposed Wasserstein-metric-based distributionally robust optimization (**WDRO**) method, three other methods are introduced for benchmarking. The first one is the robust optimization (**RO**) approach which requires the security constraints (4-15g)∼(4-15j) to be satisfied for all realizations in the support of the random variable. The second benchmarking approach is the moment-based distributionally robust optimization (**MDRO**). The ambiguity set for MDRO is the set of all probability distributions with given mean and covariance (usually sample mean and sample covariance). Then the chance constraints (4-15g)∼(4-15j) can be reformulated as SOCP constraints by leveraging the Chebyshev inequality $\mathbb{P}\{|\tilde{\xi} - \mu| \geq \sqrt{1/\rho}\sigma\} \leq \rho$. Moreover, the third approach is the Gauss-based stochastic programming (**GSP**) which presumes that the random variable follows Gauss distribution with given mean and covariance. Then the chance constraints (4-15g)∼(4-15j) can also be reformulated into SOCP constraints by using the inequality $\mathbb{P}\{|\tilde{\xi} - \mu| \geq \Phi^{-1}(1-\rho/2)\sigma\} \leq \rho$ where Φ is the cumulative distribution function of standard Gauss random variable. In all three benchmarking approaches described above, the expectation in the objective function (4-14) is estimated by the sample average of the cost function. It seems that the MDRO and GSP cannot be integrated into the extensible OPF structure provided by MATPOWER due to the introduction of SOCP constraints. However, rather than directly working with the SOCP constraints, the cutting-plane approach illustrated in [41] and [43] iteratively employs a sequence of linear cuts to the SOCP constraints. In each iteration, only linear constraints are involved in enforcing the chance constraints (4-15g)∼(4-15j), therefore it fits well in the extensible OPF structure provided by MATPOWER.

To test and compare the performance of different methods to deal with uncertainties, an underlying random number generator (RNG) is employed to simulate the VRE forecasting errors. Following the suggestion in [120], we use Laplace distribution to generate "realistic" historical data for wind power forecasting errors with the typical standard error in U.S. reported

in [85]. The parameters of RNG including the type of distribution is secret and only the data generated from the RNG is available to all the methods under test. From a finite number of data, each method constructs its own optimal operation strategy. Then the RNG is again used to generate a much larger set of data for Monte Carlo simulation (MCS) to assess the statistical performance of the operation strategy obtained by each method.

4.5.2 IEEE 14-Bus System

The diagram of the modified IEEE 14-bus system is shown in Fig. 4-4. Four wind farms with each capacity of 36 MW are installed at bus 11, 12, 13 and 14. The forecasting values of wind power are 50% of their capacity. The transmission capacity of each line is set to be 40 MW. The regulating reserve prices are assumed to be 50% of the linear coefficients of the generator cost functions. The following group of tests is designed to demonstrate the features of the proposed approximate AC-OPF model.

Table 4-1 presents the optimal operation strategy of the proposed model under different conditions with the comparison to the DC model. When operated with the strategy given by the complete model, all the chance constraints for reserve, voltage, reactive power and line flow are well guaranteed. For example, Fig. 4-5 shows the normalized histogram of system reserve usage, voltage magnitude at bus 12, reactive power output of G1 and line flow in line 4-5. It is shown those quantities safely distributed in the allowable ranges. If we exclude the voltage chance constraints (4-15h) from the proposed model, the corresponding optimal generation strategy is also given in Table 4-1. Compared with the strategy of complete model, the nominal active power and participation factors of generators do not have many differences, but the setting voltage magnitudes at generator buses increase significantly. The consequence of this change is shown in Fig. 4-6 (a). Under the voltage unconcerned strategy, the bus 12 will be exposed to the huge risk of over-voltage. On the other hand, if we drop the generator reactive power limits (4-15i) from the complete model, the obtained operation strategy is also presented in Table 4-1. Again, the nominal active power and participation factors only exhibit slight changes whereas the AVR setting points vary significantly. For example, the setting voltage of G1 decreases from 1.036 to 1.029. The consequence this change is illustrated in Fig. 4-6 (b) which shows G1 is under significant risk of crossing the under-excitation limit. Above discussion indicates that both voltage chance constraint (4-15h) and the reactive power chance constraint (4-15i) are essential for a safe and meaningful operation strategy under uncertainties. Table 4-1 further lists the strategy obtained by the DC model widely used in the literature [41–43, 107, 108]. Except for the voltage and reactive power unawareness of DC model, more importantly, the nominal active power and participation factors exhibit non-negligible differences compared with those of the proposed model. This shows that a more accurate power flow model could significantly improve the operation strategy obtained by DC model.

So it is important to ask how accurate is the approximate AC-OPF model in this chapter? In our model (4-15), we have an exact AC model (4-1) to govern the nominal operation point and the approximate LPF model (4-5)(4-6) to calculate the incremental response under uncertainties. The accuracy, therefore, should lie between the LPF model (4-3) and the exact AC model. The following group of tests is conducted under the operation strategy obtained by the proposed complete model shown in Table 4-1. Table 4-2 shows the operation costs calculated using different models as the total VRE forecasting error varies from -32 MW to +32 MW. When the forecasting error is zero, our model coincides with the full-AC model, so the error of operation costs is zero. For other values of total forecasting error, the approximate AC model in this chapter shows much higher accuracy than the LPF model in calculating the operation costs. Table 4-3 and Table 4-4 further show the accuracy of calculating voltage magnitudes and reactive power outputs when the total VRE forecasting error is -32 MW. It is shown the accuracy of the approximate AC model in this chapter is at least one-order-of-magnitude higher than that of the LPF model for calculating both voltage and reactive power. Since the approximate AC model in this chapter inherits the formula (4-4) of the LPF model, it has the same accuracy as the LPF model in calculating line MW flow but has higher accuracy than the DC model as shown in Table 4-5. In summary, the proposed approximate AC model combines the accuracy of the full-AC model and the tractability of LPF model, which makes it a more attractive model for OPF under uncertainties.

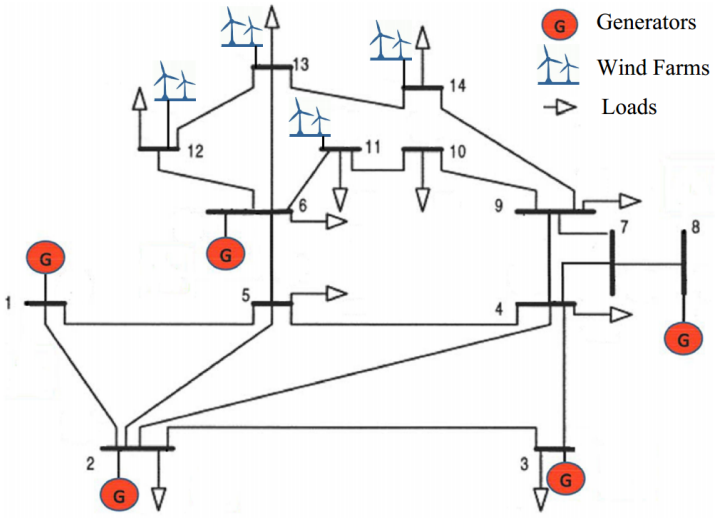


Figure 4-4 Diagram of the Modified IEEE 14-bus System

4.5.3 IEEE 118-Bus System

On the IEEE 118-bus system, 18 wind farms are installed at bus 2, 5, 7, 13, 15, 21, 25, 28, 35, 45, 53, 58, 63, 75, 88, 95, 106 and 115, respectively. The capacity of each wind farm changes from 30 to 90 MW to test the methods under different levels of wind penetration and

Table 4-1 Optimal Operation Strategy of the Proposed Model under Different Conditions with Comparison to the DC Model

	the proposed complete model			without voltage constratints			without Qg constratints			DC Model		
	Pg	alpha	Voltage	Pg	alpha	Voltage	Pg	alpha	Voltage	Pg	alpha	Voltage
G1	0.562	0.061	1.036	0.561	0.061	1.058	0.566	0.066	1.029	0.555	0.062	/
G2	0.339	0.292	1.026	0.339	0.293	1.048	0.343	0.309	1.019	0.338	0.195	/
G3	0.538	0.000	1.024	0.539	0.000	1.045	0.546	0.000	1.017	0.454	0.000	/
G4	0.376	0.535	1.032	0.376	0.535	1.055	0.370	0.525	1.033	0.386	0.548	/
G5	0.079	0.112	0.995	0.078	0.111	1.018	0.070	0.100	0.986	0.137	0.195	/

Table 4-2 Operation Costs ($) Calculated by Different Models

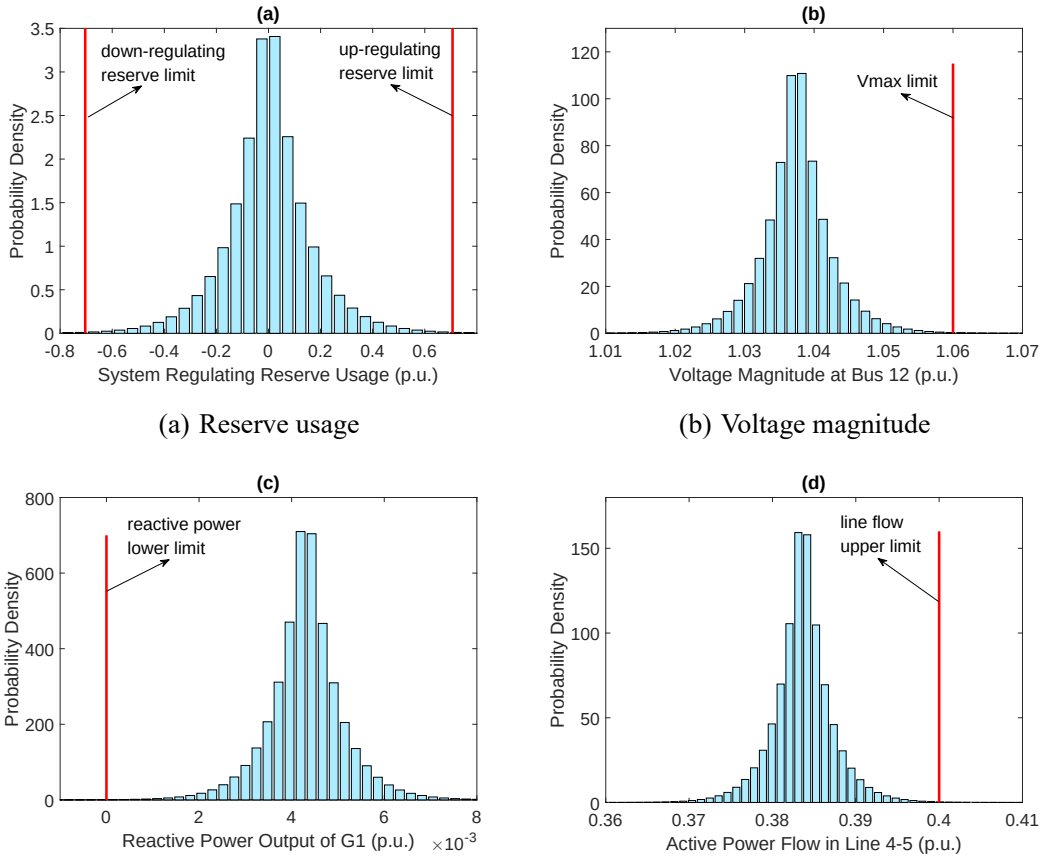
VRE forecast error	full-AC	this chapter	**error(%)**	LPF	**error(%)**
-32	7502.07	7500.66	**0.02%**	7440.90	**0.82%**
-24	7180.80	7180.89	**0.00%**	7121.23	**0.83%**
-16	6863.44	6864.26	**-0.01%**	6804.70	**0.86%**
-8	6549.98	6550.76	**-0.01%**	6491.31	**0.90%**
0	6240.41	6240.41	**0.00%**	6181.06	**0.95%**
8	5934.71	5933.19	**0.03%**	5873.94	**1.02%**
16	5632.86	5629.11	**0.07%**	5569.96	**1.12%**
24	5334.86	5328.17	**0.13%**	5269.12	**1.23%**
32	5040.69	5030.36	**0.21%**	4971.41	**1.37%**

Table 4-3 Voltage Magnitudes (p.u.) Calculated by Different Models

Bus No.	full-AC	this chapter	**error**	LPF	**error**
4	1.019	1.019	**-0.0001**	1.017	**-0.0023**
5	1.025	1.025	**-0.0001**	1.024	**-0.0012**
7	1.002	1.002	**-0.0001**	0.993	**-0.0094**
9	0.998	0.998	**-0.0002**	0.979	**-0.0192**
10	0.997	0.997	**-0.0003**	0.982	**-0.0153**
11	1.015	1.014	**-0.0003**	1.007	**-0.0073**
12	1.028	1.028	**-0.0004**	1.027	**-0.0012**
13	1.018	1.017	**-0.0001**	1.015	**-0.0025**
14	0.995	0.995	**-0.0003**	0.984	**-0.0114**

Table 4-4 Reactive Power Output (p.u.) Calculated by Different Models

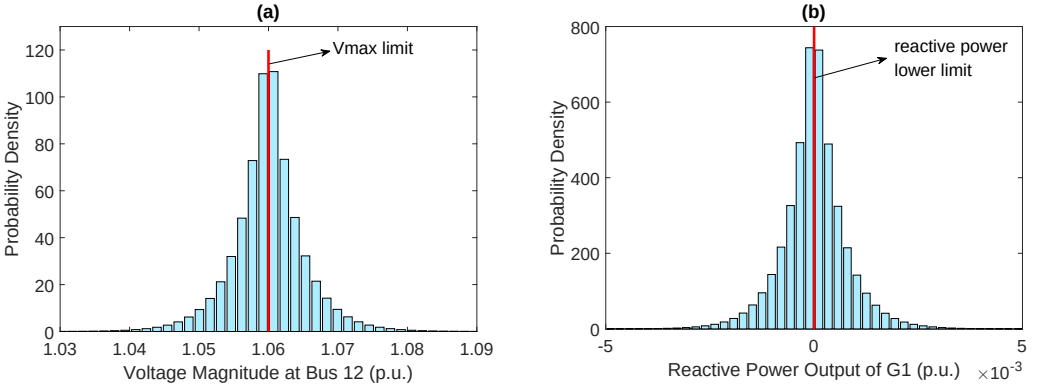
Unit No.	full-AC	this chapter	**error**	LPF	**error**
G1	0.006	0.006	**-0.0006**	0.005	**-0.0011**
G2	-0.101	-0.103	**-0.0023**	-0.094	**0.0067**
G3	0.298	0.297	**-0.0005**	0.301	**0.0034**
G4	0.211	0.213	**0.0025**	0.259	**0.0480**
G5	-0.041	-0.041	**-0.0001**	0.011	**0.0519**



(a) Reserve usage (b) Voltage magnitude

(c) Reactive power (d) Line flow

Figure 4-5 Normalized histogram of system reserve usage, voltage magnitude at bus 12, reactive power output of G1 and line flow in line 4-5 by MCS when operated at the strategy given by the proposed complete model



(a) Voltage magnitude (b) Reactive power

Figure 4-6 Normalized histogram of voltage magnitude at bus 12 (a) for the proposed model without voltage constraint and the normailized histogram of reactive power output at G1 (b) for the proposed model without reactive power constraint.

uncertainties.

Table 4-5 Line MW Flow (p.u.) Calculated by Different Models

line	full-AC	this chapter	**error**	DC	**error**
4-5	-0.405	-0.390	**0.0149**	-0.340	**0.0650**
1-2	0.388	0.374	**-0.0139**	0.323	**-0.0651**
2-3	0.292	0.291	**-0.0019**	0.261	**-0.0315**
2-4	0.199	0.195	**-0.0035**	0.148	**-0.0509**
10-11	-0.199	-0.204	**-0.0058**	-0.185	**0.0133**

In our implementation, the tolerable violation probability in (4-15g)∼(4-15j) is set to $\rho_1 = \rho_2 = \rho_3 = \rho_4 = 0.05$, and the confidence level in (4-24) is set to 0.9. The forecasting values of wind generation are set to be 50% of the installed capacity. The available sample size ranges from 10^2 to 10^6 to showcase the data-exploiting feature. After solving each problem, Monte Carlo simulation with 10^7 samples is employed to test the practical and out-of-sample performance. The major test results are summarized in Table 4-6. When the capacity of each wind farm is 30MW, all the tested methods obtain the corresponding optimal solutions. The evolution of the operational costs by Monte Carlo simulation is illustrated in Fig. 4-7. The costs of WDRO and MDRO lie between RO and GSP due to the fact that RO completely ignores the probabilistic information whereas the GSP assumes precise knowledge about the probability distribution. In other words, RO and GSP produce the most conservative and aggressive strategies, respectively. MDRO assumes partial knowledge, the first and second moments, of the probability distribution, which reduces conservatism compared with RO to some degree. In contrast, the proposed WDRO fully relies on the information told by the data at hand. When we are in short of data, it approaches the RO method to take a conservative decision. On the contrary, when the data is rich, it approaches the stochastic programming approach with complete information about the distribution. Fig. 4-8 further compares the lowest reliability level of the security constraints. Although GSP obtains the strategy with lowest operation costs, it fails to guarantee the required reliability level. This is due to the deviation of the underlying true distribution from the Gauss assumption made in GSP method. All other methods, including RO, MDRO and the proposed WDRO ensure higher reliability level than required due to their "robust" nature. As more data is available, the proposed WDRO gradually and safely reduce the guaranteed reliability level to pursue higher economic efficiency. Note that the proposed WDRO considers the expected costs in the objective function (4-14) w.r.t. the worst-case distribution and the true distribution usually differs from the worst-case one, so the simulated costs are always upper bounded by the objective function as shown in Table 4-6. Fig. 4-9 shows the evolution of percentage difference between the objective function and simulated costs for the proposed WDRO. As more data is available, the ambiguity set shrinks and worst-case distribution approaches the true distribution, and the difference between simulated

costs and objective function diminishes. As shown in Table 4-6, when the installed capacity of wind power increases to 60×18 MW, the RO along with the WDRO with 10^2 and 10^3 samples become infeasible. When the installed capacity increases to 90×18 MW the MDRO and the WDRO with 10^4 also become infeasible. In summary, we can order the methods according to their conservatism as: RO>WDRO($10^{2\sim3}$)>MDRO>WDRO($10^{4\sim6}$)>GSP.

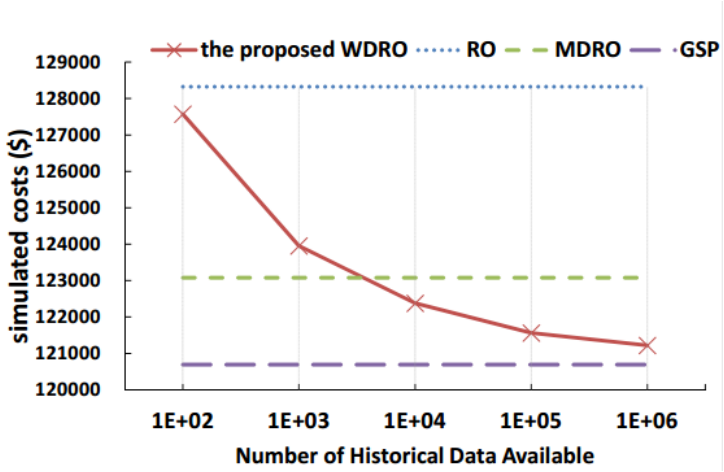


Figure 4-7 Evolution of simulated operational costs as more historical data is available with 30×18 MW wind integration.

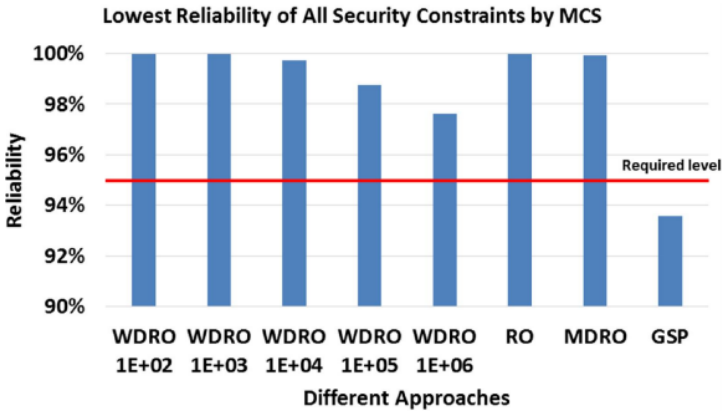


Figure 4-8 Comparison of the lowest reliability of all security constraints by Monte Carlo simulation

The computation of the proposed method consists of two stages. One is the preparation stage to construct the uncertainty set $\mathcal{U}$ in (4-34), and the other is the operation stage to solve the optimization problem (4-14)(4-15). In practice, the preparation stage can be done off-line and updated on weekly or monthly basis according to the new data available. In addition, the construction of the uncertainty set is highly parallelizable among different buses and transmission lines. The time on our 12-core workstation to construct uncertainty set for the whole 118-bus system is listed in Table 4-7. The computation time grows linearly with the sample size N, and further speed-up is possible by leveraging distributed computation. The computation time for the on-line operation stage is listed in Table 4-6. In this stage, the solver time is nearly

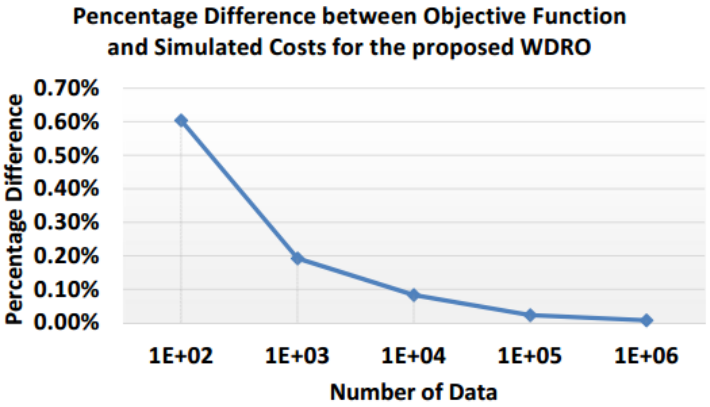


Figure 4-9 The procedure to determine the uncertainty set $\mathcal{U}$.

Table 4-6 Comparative Case Study on IEEE 118-bus System

Wind Capacity	Methods / Performance	the proposed WDRO with different sample sizes					RO	MDRO	GSP
		1.E+02	1.E+03	1.E+04	1.E+05	1.E+06			
30*18 MW	objective ($)	128345	124192	122483	121588	121227	128355	123168	120839
	simulated costs ($)	127574	123953	122381	121560	121218	128322	123081	120694
	up reverse (p.u.)	4.296	2.504	1.712	1.302	1.130	4.672	2.038	0.828
	down reverse (p.u.)	4.319	2.490	1.711	1.299	1.129	4.690	2.084	0.906
	reliability	99.99997%	99.98933%	99.74943%	98.75797%	97.61366%	100.00000%	99.93666%	93.56586%
	cpu time (s)	0.31	0.27	0.31	0.23	0.29	0.45	0.27	0.29
60*18 MW	objective ($)	Infeasible	Infeasible	115404	113883	113017	Infeasible	116608	112015
	simulated costs ($)			115197	113814	112990		116692	111867
	up reverse (p.u.)			3.340	2.672	2.258		4.128	1.657
	down reverse (p.u.)			3.344	2.677	2.259		4.081	1.738
	reliability			99.70579%	98.90693%	97.61284%		99.93669%	93.18895%
	cpu time (s)			0.32	0.44	0.27		1.14	0.26
90*18 MW	objective ($)	Infeasible	Infeasible	Infeasible	106522	105369	Infeasible	Infeasible	103719
	simulated cost ($)				106391	105315			103582
	up reverse (p.u.)				3.877	3.410			2.575
	down reverse (p.u.)				3.890	3.407			2.654
	reliability				98.72546%	97.66914%			93.75508%
	cpu time (s)				0.33	0.32			1.72

the same as the RO approach and less than that of MDRO and GSP methods. More importantly, the solver time of WDRO approach at the operation stage is irrelevant to the number of wind farms and amount of historical data at hand. Therefore, the proposed WDRO approach is highly scalable and efficient for real-time operation.

4.6 Conclusion

This chapter has proposed a chance-constrained approximate AC-OPF under uncertainties based on distributionally robust optimization with Wasserstein metric. In order to overcome the flaws of the DC power flow model extensively used in stochastic OPF formulations, we have

Table 4-7 Time (s) to Construct Uncertainty Sets for 118-bus System

N	1.E+02	1.E+03	1.E+04	1.E+05	1.E+06
Time (s)	0.67	0.82	1.64	16.48	291.64

developed a tractable AC power flow model by integrating the full AC power flow and a linear power flow models. Numerical studies have demonstrated the proposed OPF formulation has improved precision as well as good numerical tractability. Without any assumption on the underlying probability distribution of the uncertainties, the proposed chance-constrained AC-OPF is completely driven by the available historical data. It extracts reliable probabilistic information from historical data to construct an ambiguity set of all possible distributions and immunizes the operation strategy against all distributions in the ambiguity set. The more data is available, the less conservative the solution is. In addition, special problem structures are properly exploited in the reformulation of the distributionally robust optimization problem to improve the scalability and efficiency of the numerical solution approach.

5 Steady-state Optimal Allocation of FACTS Devices via Sparse Optimization

5.1 Introduction

FACTS devices can be utilized to increase transmission capability and improve stability in modern power systems. In order to maximize the benefits of installing FACTS devices, their types, location, capacity, and even initial settings, should be systematically determined. Usually this problem is called the FACTS devices allocation problem and has attracted much attention in the past two decades [121–125]. Various formulations and algorithms have been put forward to deal with this problem. Different objective functions are proposed from the perspectives of system economy and/or security, such as investment costs [121, 122], transmission losses [123], generator fuel costs [124], voltage stability index [125], voltage profile [122] and system loadability [126].

In fact, this problem is theoretically a mixed integer nonlinear programming (MINP) problem, and still there are no general and effective mathematical techniques to solve such problem, especially when the scale of the problem is large. To handle the difficulty of MINP, sensitivity analysis [127], intelligent optimization algorithms [128, 129] and mixed integer linear programming (MILP) have been extensively investigated in previous literatures [121, 130, 131].

The basic idea of sensitivity analysis methods is to find some indicators to determine the most critical transmission lines or buses for installing the FACTS devices [127]. Sensitivity analysis methods have their advantages over other optimization-based methods in computing efficiency. However, their computation accuracy is partly lost, because the nonlinearity of the power flow model is neglected. Moreover, they can not simultaneously optimize the device number, location as well as initial settings.

Intelligent optimization algorithms have their merits in dealing with discrete variables and finding the global optimal solution, so they have been widely applied to FACTS devices allocation problems [128, 129]. Unfortunately the demerits of this type of algorithms are their high computational burden.

MILP based algorithms either relax or approximate the original nonlinear formulation to a linear model and they can be classified into two groups: the relaxation group and the approximation group. In the relaxation group, decomposition techniques are employed successively to reformulate the original MINP into MILP [121, 130, 131]. However, recursively solving MILP is too time-consuming for large-scale systems. In the approximation group, the nonlinear power flow is approximated to simpler models such as the DC model [132] and the simplified LFB (line flowed based) model [122]. Though algorithms in the approximation group have

relatively high computational efficiency, the approximation nature of the problem formulation makes those algorithms merely suitable for initial analysis in power system planning, and their results need to be further refined by a full AC power flow model.

Problems in various science and engineering fields motivate the need for a sparse solution [133]. In 2012, R. A. Jabr et al. applied sparse optimization for the first time in power system VAR planning [134] , where L_1 regularization was combined with successive conic programming to achieve a sparse solution vector, i.e. a VAR allocation strategy with minimum installation sites.

In fact, FACTS devices allocation problems possess following features: 1) large-scale mixed integer nonlinear programming problems; 2) a large number of transmission lines and buses which implies a large quantity of potential locations to install FACTS devices; 3) limited actual number of FACTS devices to be installed due to economical consideration. Considering those features, in this chapter, FACTS devices allocation problems are considered as a sparse optimization problem, by introducing an extra constraint, i.e. the solution vectors must be sparse [135].

A new algorithm is proposed to solve these sparse optimization problems, which consists of the following four parts: 1) sparsity-inducing norms [133], 2) alternating direction method of multipliers (ADMM) [136], 3) interior point method (IPM) [137, 138] and 4) shrinkage-thresholding operators (STO) [139–141], and named as ADMM-IPM-STO. The objective function (system loadability) is regularized by several sparsity-inducing norms. $L_{1/2}$ and $L_{2/3}$, recently developed in [140] and [141], are firstly introduced in power system optimization in this chapter. Additionally, L_1 norm is also employed in the numerical experiments. The features of different sparsity-inducing norms is discussed. The sparse optimization problem is decomposed into two sub-problems by ADMM, namely a nonlinear programming (NLP) sub-problem and an L_q regularization sub-problem. Then the NLP sub-problem is solved by IPM, while the L_q regularization sub-problem has a closed-form analytic solution expressed by STO. The theoretical convergence of the proposed method is analyzed. A weak result that guarantees the optimality of the practical solutions of all $q \in (0, 1]$ and a strong result that substantiates the convergence for $q = 1$ are obtained.

Within above framework, the state-of-art method for sparse optimization is combined with one of the most successful methods for optimal power flow to form a novel FACTS device allocation algorithm. The numbers, locations, initial settings and even types of FACTS devices are simultaneously determined by this algorithm. Its computation precision is relatively higher than that of DC or LFB based algorithms, because the exact AC flow model is adopted. On the other hand, the proposed algorithm has relatively higher computational efficiency than intelligence optimization algorithms which also use AC flow model. Furthermore, this algorithm is adaptive to various types of FACTS allocation problems and possesses the flexibility to use

different sparsity-inducing norms to achieve desirable sparse features.

Case studies are carried out on IEEE test systems from 9 buses to 300 buses, respectively, which demonstrates validity and above mentioned advantages of the proposed algorithm.

The rest of this chapter is organized as follows. Notations used in this chapter are defined in section 5.2. The mathematical formulation of general sparse optimization problem and the ADMM-IPM-STO algorithm, along with the convergence analysis of ADMM-IPM-STO are introduced in section 5.3. The sparse optimization model for general FACTS devices allocation problems is presented in section 5.4. Case studies are shown in section 5.5. The results of case studies are analyzed and discussed in section 5.6. Finally, section 5.7 draws conclusions and discuss extensions of the proposed formulation and algorithm.

Nomenclature

The imaginary unit is denoted by j. Boldface lower case letter $\boldsymbol{a}$ represents a real vector and its i-th element is denoted by a_i. Hatted boldface lower case letter $\hat{\boldsymbol{b}}$ represents a complex vector with its i-th element denoted by $\hat{b}_i$. The set of real and complex n-vectors are denoted by $\mathbf{R}^n$ and $\mathbf{C}^n$. The conjugate of a complex number $\hat{b}$ is denoted by $\hat{b}^c$. $\mathrm{Re}(\hat{b})$ and $\mathrm{Im}(\hat{b})$ denote the real and imaginary part of $\hat{b}$. We denote the gradient of the function $f(\boldsymbol{x})$ at the point $\boldsymbol{x}^*$ as $\nabla f(\boldsymbol{x}^*)$. The first-order directional derivative of $f(\boldsymbol{x})$ at the point $\boldsymbol{x}^*$ toward direction $\boldsymbol{d}$ is denoted by $f'(\boldsymbol{x}^*; \boldsymbol{d})$. $T_S(\boldsymbol{x}^*)$ denotes the tangent cone of set S at point $\boldsymbol{x}^*$. Vector and scalar sequences are denoted with superscript like $\boldsymbol{x}^k$ and α^k. The δ-neighborhood of vector $\boldsymbol{x} \in \mathbf{R}^n$ is denoted as $B(\boldsymbol{x}, \delta) = \{\boldsymbol{y} \in \mathbf{R}^n | \ \|\boldsymbol{y} - \boldsymbol{x}\|_2 < \delta\}$. We use parentheses to construct vectors from comma separated lists as $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = [\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_k^T]^T$.

In addition, the following special symbols are used in our problem formulation:

n_b	number of buses
n_g	number of generators
n_l	number of lines
$\hat{y}$	series admittance of a line
jb	shunt admittance of a line
$\hat{v}$	complex voltage at a bus
e	real part of the voltage at a bus
f	imaginary part of the voltage at a bus
$\hat{i}_g$	complex current injection of a generator
a_g	real part of the current injection of a generator
b_g	imaginary part of the current injection of a generator
b_{sh}	susceptance of a shunt compensator
κ	compensation rate of a series compensator

φ angle of a phase shifter

τ ratio of a controllable transformer

5.2 Formulation and Algorithm for General Sparse Optimization

5.2.1 Sparse Optimization Problems

A standard nonlinear optimization problem usually is expressed as

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in S} f(\boldsymbol{x},\boldsymbol{u}) \tag{5-1}$$

where $f(\cdot)$ is a nonlinear scalar objective function, $x \in \mathbf{R}^m$ and $u \in \mathbf{R}^n$ are two sub-decision vectors. S is the feasible set of the decision variables, and it will be specified by equality and inequality constraints later on.

In some applications, e.g. FACTS devices allocation problems, decision vector $\boldsymbol{u}$ is expected to be very sparse after optimization. This expectation can be considered an extra constraint in above original optimization problem (5-1). This kind of problem is named the sparsity-constrained optimization problem which is intuitively transformed into the so-called L_0 regularization problem [142]:

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in S} f(\boldsymbol{x},\boldsymbol{u}) + \lambda\|\boldsymbol{u}\|_0 \tag{5-2}$$

where $\|\cdot\|_0$, called L_0 norm, is the number of nonzero components of $\boldsymbol{u}$. The non-negative parameter λ, given by decision-makers, balances the two objective terms. Obviously, the larger λ is, the sparser induced $\boldsymbol{u}$ will be. However, L_0 regularization problem is NP hard [143] due to the discrete nature of L_0 norm. Actually, it can be seen from (5-4) that discrete L_0 norm is the limit of the q power of continuous L_q norm when q approaches 0. A natural way to overcome the difficulty in solving the L_0 regularization problem is to consider the so-called L_q regularization problem

$$\min_{(\boldsymbol{x},\boldsymbol{u})\in S} f(\boldsymbol{x},\boldsymbol{u}) + \lambda\|\boldsymbol{u}\|_q^q \tag{5-3}$$

where

$$\|\boldsymbol{u}\|_q = (\sum_{i=1}^{n}|u_i|^q)^{1/q} \quad (0 < q \leq 1) \tag{5-4}$$

According to the definition (5-4), L_q norms, to some extent, are approximations to L_0 norm. As q approaches 0, L_q norm mainly exerts penalty on the number of nonzero components of the solution vector. Comparatively, as q reaches 1, L_q norm penalizes the sum of absolute values of the solution vector components. When q lies somewhere between 0 and 1, L_q norm exerts penalty on both the number of nonzero components and the sum of absolute values to a certain degree. In other words, for a certain λ, the sparsity of the solution vector to the L_q

regularization problem increases as q decreases. The special importance of $L_{1/2}$ regularization is highlighted in [144] by showing the representativeness of $L_{1/2}$ regularization among all L_q regularizations. This work basically reveals that the sparsity of the L_q solution significantly increases as q decreases when $1/2 < q \leq 1$ but is insignificantly affected by q when $0 < q \leq 1/2$. In addition, thresholding representation theories have been developed for L_q regularization problems when $q = 1/2$, $2/3$ and 1 [139–141], which leads to efficient algorithms for solving those L_q regularization problems. Therefore, $L_{1/2}$, $L_{2/3}$ and L_1 norms are so far the best sparsity-inducing norms to obtain desirable sparse solutions.

5.2.2 ADMM-IPM-STO Algorithm for Sparse Optimization

Consider the L_q regularization problem (5-3). The objective function consists of two terms and the second term is a continuous, non-smooth, non-lipschitz function of $\boldsymbol{u}$. Conventional joint minimization methods are incapable to tackle this problem.

ADMM is one of the state-of-art methods for sparse optimization. A comprehensive account of ADMM appears in [136] and its applications in sparse optimization are reported in [145]. ADMM is actually a version of the method of multipliers where Gauss-Seidel iterations are used to separately minimize two terms in the objective function instead of conventional joint minimization. ADMM utilizes the separability structure of the objective in (5-3) and decomposes problem (5-3) into two simpler sub-problems.

In order to make the objective separable, first an auxiliary vector $\boldsymbol{v}$ and an auxiliary equality constraint are introduced, and then problem (5-3) can be equivalently transformed into

$$\begin{aligned} &\underset{(\boldsymbol{x},\boldsymbol{u})\in S,\boldsymbol{v}\in\mathbf{R}^n}{min} f(\boldsymbol{x},\boldsymbol{u}) + \lambda\|\boldsymbol{v}\|_q^q \\ &s.t. \quad \boldsymbol{u} - \boldsymbol{v} = \mathbf{0} \end{aligned} \tag{5-5}$$

The augmented Lagrangian function with respect to the auxiliary equality constraint is given by

$$L_\rho(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{y}) = f(\boldsymbol{x},\boldsymbol{u}) + \lambda\|\boldsymbol{v}\|_q^q + \boldsymbol{y}^T(\boldsymbol{u}-\boldsymbol{v}) + \rho/2\|\boldsymbol{u}-\boldsymbol{v}\|_2^2 \tag{5-6}$$

where $\boldsymbol{y}$ is an n dimensional Lagrangian multiplier vector related to the auxiliary equality constraint, ρ is a positive penalty parameter.

ADMM-IPM-STO algorithm consists of the following iterations:

$$(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1}) := arg\underset{(\boldsymbol{x},\boldsymbol{u})\in S}{min} L_\rho(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}^k,\boldsymbol{y}^k) \tag{5-7}$$

$$\boldsymbol{v}^{k+1} := arg\underset{\boldsymbol{v}\in\mathbf{R}^n}{min} L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v},\boldsymbol{y}^k) \tag{5-8}$$

$$\boldsymbol{y}^{k+1} := \boldsymbol{y}^k + \rho(\boldsymbol{u}^{k+1} - \boldsymbol{v}^{k+1}) \tag{5-9}$$

Note that problem (5-7) is a conventional continuous nonlinear optimization problem whose decision vectors are only $\boldsymbol{x}$ and $\boldsymbol{u}$. Various methods have been proposed to solve this NLP problem, among which IPM has experienced great success. Specially, IPM has almost become a standard method to solve optimal power flow problems in recent years. Hence, IPM is chosen to solve the NLP sub-problem. When problem (5-7) solved, $\boldsymbol{x}^{k+1}$ and $\boldsymbol{u}^{k+1}$ are obtained. Problem (5-8) thus can be equivalently stated as

$$\boldsymbol{v}^{k+1} := \underset{\boldsymbol{v}\in\mathbf{R}^n}{argmin}(\lambda/\rho)\|\boldsymbol{v}\|_q^q + (1/2)\|\boldsymbol{u}^{k+1} - \boldsymbol{v} + (1/\rho)\boldsymbol{y}^k\|_2^2 \tag{5-10}$$

So far, there is no general theoretical understanding and efficient algorithms to arbitrary $q \in (0, 1]$, because L_q norm in (5-10) is non-convex, nonsmooth and non-Lipschitz. Fortunately, at special points of $q = 1/2$, $2/3$, and 1, closed-form analytic solutions have been established in [139–141] which give global optimum for (5-10). They all can be expressed as shrinkage-thresholding operators. Though the exact forms vary with different sparsity-inducing norms, they are uniformly given by

$$\boldsymbol{v}^{k+1} = \text{Shrink}\left(\boldsymbol{u}^{k+1} + (1/\rho)\boldsymbol{y}^k, \lambda/\rho\right) \tag{5-11}$$

The specific expression of STOs for $L_{1/2}$ [140], $L_{2/3}$ [139] and L_1 [141] norms are presented here. They are all diagonally nonlinear operators uniformly expressed as

$$\text{Shrink}(\boldsymbol{z}, \beta) = [H_\beta(z_1) \quad H_\beta(z_2)\ldots \quad H_\beta(z_n)]^T \tag{5-12}$$

For $q = 1/2$,

$$H_\beta(z_i) = \begin{cases} \frac{2}{3}|z_i|\left(1 + \cos\left(\frac{2\pi}{3} - \frac{2\varphi_\beta(z_i)}{3}\right)\right) & z_i > p(\lambda) \\ 0 & |z_i| \le p(\lambda) \\ -\frac{2}{3}|z_i|\left(1 + \cos\left(\frac{2\pi}{3} - \frac{2\varphi_\beta(z_i)}{3}\right)\right) & z_i < -p(\lambda) \end{cases} \tag{5-13}$$

where $\varphi_\beta(z_i) = \arccos\left(\frac{\beta}{8}\left(\frac{|z_i|}{3}\right)^{-\frac{2}{3}}\right)$ and $p(\lambda) = \frac{\sqrt[3]{54}}{4}(\beta)^{\frac{2}{3}}$.

For $q = 2/3$,

$$H_\beta(z_i) = \begin{cases} \left(\frac{m+\sqrt{2|z_i|/m - m^2}}{2}\right)^3 & z_i > p(\beta) \\ 0 & |z_i| \le p(\beta) \\ -\left(\frac{m+\sqrt{2|z_i|/m - m^2}}{2}\right)^3 & z_i < -p(\beta) \end{cases} \tag{5-14}$$

where $m = \frac{2}{\sqrt{3}}\beta^{\frac{1}{4}}\left(\cosh\left(\frac{\phi}{3}\right)\right)^{\frac{1}{2}}$, $\phi = \text{arccosh}\left(\frac{27z_i^2}{16}\beta^{-\frac{3}{2}}\right)$ and $p(\beta) = \frac{2}{3}(3\beta^3)^{\frac{1}{4}}$.

For $q = 1$,

$$H_\beta(z_i) = \text{sgn}(z_i)\max\{|z_i| - \beta, 0\} \tag{5-15}$$

The computation burden of obtaining $\boldsymbol{v}^{k+1}$ can almost be omitted since above closed-form

solutions involve no iterations. Then according to the method of multipliers, dual variable $\boldsymbol{y}^k$ needs to be updated to $\boldsymbol{y}^{k+1}$ as in (5-9). The iteration terminates when the primal error (5-16) and dual error (5-17) are both small enough.

$$\|\boldsymbol{u}^{k+1} - \boldsymbol{v}^{k+1}\|_2 < \epsilon_{primal} \tag{5-16}$$

$$\|\boldsymbol{v}^{k+1} - \boldsymbol{v}^{k}\|_2 < \epsilon_{dual} \tag{5-17}$$

5.2.3 Convergence Analysis of ADMM-IPM-STO

The convergence of ADMM is well-established for convex problems in [136]. So far, there is no global convergence result of ADMM for general non-convex optimization problems. Even so, ADMM has been extensively applied to non-convex problems [146, 147], including OPF problems [148, 149] and shown robust performance in practice. The convergence analysis of an optimization algorithm can be divided into two questions. First, whether does the algorithm generate a limit point? Second, whether is the limit point an optimum? Both questions are of crucial importance to substantiate the algorithm. In engineering application, the second question is to some extent more important than the first one because emergence of a limit point can be directly observed through numerical computation while optimality is not obviously available. Since the proposed problem formulation is non-convex, and even non-lipschitz, the above two questions are in doubt. In this part, we establish a weak result (Theorem 5.2) for problem (5-5) with $0 < q \leq 1$ to answer the second question and a strong result (Theorem 5.3) with $q = 1$ to partly answer the first one. In the weak result, we basically adopt the same approach as in [146] and [147], whereas the strong result is essentially based on [136].

Because $|x_i|^q$ regularization term in problem (5-5) is non-lipschitz at 0 when $0 < q < 1$ [140], Lagrange multipliers lose their geometric meaning and therefore KKT optimality condition in standard smooth and non-smooth optimization theories cannot apply to its analysis. But note that the first-order directional derivative of $|x_i|^q$ at 0 still exists if it is allowed to take infinite value, i.e. $\lim_{t\to 0^+} t^{q-1}|d_i|^q = +\infty$. Denote the second term in the objective function of problem (5-5) as $g(\boldsymbol{v})$. Thus the following analysis will be based on the first-order directional derivatives.

We first give a sufficient optimality condition for problem (5-5) in the context of ADMM. For simplicity, define $h(\boldsymbol{x}, \boldsymbol{u}) = f(\boldsymbol{x}, \boldsymbol{u}) + g(\boldsymbol{u})$. We say an optimization problem $\min f(\boldsymbol{x})$ s.t. $\boldsymbol{x} \in S$ is generic if the first-order necessary optimality condition $f^{'}(\overline{\boldsymbol{x}}; \boldsymbol{d}) \geq 0,\ \forall \boldsymbol{d} \in T_S(\overline{\boldsymbol{x}})$ is also sufficient for $\overline{\boldsymbol{x}}$ to be a strong local minimum. Requesting a problem to be generic is equivalent to asking that the problem is solvable in the sense of finding a local minimum by first-order optimality condition based algorithms, e.g. IPM. Since IPM is known quite robust

to solve OPF problems to local optimality and is also a building block of the proposed ADMM-IPM-STO method, it is thus necessary and reasonable to assume that the problem is generic.

We present a Lemma which will be frequently used in the subsequent discussions.

Lemma 5.1： Assume that $\overline{\boldsymbol{x}}$ is a local minimum of $min\ h_1(\boldsymbol{x}) + h_2(\boldsymbol{x}), s.t.\ \boldsymbol{x} \in S$; the first-order directional derivative $h_1'(\overline{\boldsymbol{x}}; \boldsymbol{d})$ exists; $h_2(\boldsymbol{x})$ is continuously differentiable and $\nabla h_2(\overline{\boldsymbol{x}}) = \mathbf{0}$; If problem $min\ h_1(\boldsymbol{x}),\ s.t.\ \boldsymbol{x} \in S$ is generic, then $\overline{\boldsymbol{x}}$ is also a local minimum of $h_1(\boldsymbol{x}),\ s.t.\ \boldsymbol{x} \in S$. □

Proof： Since $\overline{\boldsymbol{x}}$ is a local minimum of $h_1(\boldsymbol{x}) + h_2(\boldsymbol{x}),\ s.t.\ \boldsymbol{x} \in S$, we have $h_1'(\overline{\boldsymbol{x}}; \boldsymbol{d}) + \nabla h_2(\overline{\boldsymbol{x}})^T d \geq \mathbf{0}, \forall d \in T_S(\overline{\boldsymbol{x}})$ where $T_S(\overline{\boldsymbol{x}})$ denotes the tangent cone of S at $\overline{\boldsymbol{x}}$. It follows from $\nabla h_2(\overline{\boldsymbol{x}}) = \mathbf{0}$ that $h_1'(\overline{\boldsymbol{x}}; \boldsymbol{d}) \geq \mathbf{0}, \forall d \in T_S(\overline{\boldsymbol{x}})$. Because problem $min\ h_1(\boldsymbol{x}),\ s.t.\ \boldsymbol{x} \in S$ is generic, $\overline{\boldsymbol{x}}$ is also a local minimum of $h_1(\boldsymbol{x}),\ s.t.\ \boldsymbol{x} \in S$. ■

Theorem 5.1： $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \overline{\boldsymbol{v}})$ is a local minimum of problem (5-5) if given $\rho > 0$, there exists $\overline{\boldsymbol{y}}$, such that ① $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}})$ is a local minimum of $L_\rho(\boldsymbol{x}, \boldsymbol{u}, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}})$; ② $\overline{\boldsymbol{v}}$ is a global minimum of $L_\rho(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \boldsymbol{v}, \overline{\boldsymbol{y}})$; ③ $\overline{\boldsymbol{u}} - \overline{\boldsymbol{v}} = \mathbf{0}$; ④ problem (5-3) is generic. □

Proof： Under condition ① and ②, considering condition ③, there exists $\delta > 0, \forall (\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}) \in F \cap B\left((\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \overline{\boldsymbol{v}}), \delta\right)$ such that $L_\rho(\boldsymbol{x}, \boldsymbol{u}, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}}) \geq f(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}})$ and $L_\rho(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \boldsymbol{v}, \overline{\boldsymbol{y}}) \geq g(\overline{\boldsymbol{v}})$. Adding the two inequalities and again noticing condition ③, we have $h(\boldsymbol{x}, \boldsymbol{u}) + \rho\|\boldsymbol{u} - \overline{\boldsymbol{u}}\|_2^2 \geq h(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}})$, i.e. $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}})$ is a local minimum of $h(\boldsymbol{x}, \boldsymbol{u}) + \rho\|\boldsymbol{u} - \overline{\boldsymbol{u}}\|_2^2$. Considering assumption ④ and using Lemma 5.1, $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}})$ is also a local minimum of $h(\boldsymbol{x}, \boldsymbol{u})$. ■

Remark 5.1： Theorem 5.1 acts as the first-order sufficient optimality condition in the context of ADMM. It allows us to analyze the algorithm without considering the detailed optimality conditions of two sub-problems. The following convergence analysis will be based on this optimality condition. □

Then we present a weak result for ADMM-IPM-STO convergence with $0 < q \leq 1$ in the following theorem.

Theorem 5.2： Let $\{(\boldsymbol{x}^k, \boldsymbol{u}^k, \boldsymbol{v}^k, \boldsymbol{y}^k)\}$ be a sequence generated by ADMM-IPM-STO. Assume that ① problem (5-3) is generic; ② the sequence $\{\boldsymbol{y}^k\}$ converges to a point, i.e. $\lim_{k \to +\infty} \boldsymbol{y}^k = \overline{\boldsymbol{y}}$. Then $\{(\boldsymbol{x}^k, \boldsymbol{u}^k, \boldsymbol{v}^k)\}$ converge to a limit point $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \overline{\boldsymbol{v}})$ which is a local minimum of problem (5-5). □

Proof： Note that $(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1})$ is a strong local minimum of $L_\rho(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}^k, \boldsymbol{y}^k)$ due to the existence of the quadratic term and assumption ①. Therefore there exists $\alpha > 0$ such that $L_\rho(\boldsymbol{x}^k, \boldsymbol{u}^k, \boldsymbol{v}^k, \boldsymbol{y}^k) - L_\rho(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1}, \boldsymbol{v}^k, \boldsymbol{y}^k) \geq \alpha\left(\|\boldsymbol{x}^k - \boldsymbol{x}^{k+1}\|_2^2 + \|\boldsymbol{u}^k - \boldsymbol{u}^{k+1}\|_2^2\right)$. Since $\boldsymbol{v}^{k+1}$ is the global minimum of $L_\rho(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1}, \boldsymbol{v}, \boldsymbol{y}^k)$, in the same way, there exists $\beta > 0$

such that $L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k},\boldsymbol{y}^{k})-L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k+1},\boldsymbol{y}^{k})\geq\beta\|\boldsymbol{v}^{k}-\boldsymbol{v}^{k+1}\|_2^2$. Considering above two inequalities and noticing (5-9) and denoting $c=\min\{\alpha,\beta\}$, we have

$$\begin{aligned}
&L_\rho(\boldsymbol{x}^{k},\boldsymbol{u}^{k},\boldsymbol{v}^{k},\boldsymbol{y}^{k})-L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k+1},\boldsymbol{y}^{k+1})\\
=&L_\rho(\boldsymbol{x}^{k},\boldsymbol{u}^{k},\boldsymbol{v}^{k},\boldsymbol{y}^{k})-L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k},\boldsymbol{y}^{k})+\\
&L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k},\boldsymbol{y}^{k})-L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k+1},\boldsymbol{y}^{k})+\\
&L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k+1},\boldsymbol{y}^{k})-L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k+1},\boldsymbol{y}^{k+1})\\
\geq&c\left(\|\boldsymbol{x}^{k}-\boldsymbol{x}^{k+1}\|_2^2+\|\boldsymbol{u}^{k}-\boldsymbol{u}^{k+1}\|_2^2+\|\boldsymbol{v}^{k}-\boldsymbol{v}^{k+1}\|_2^2\right)\\
&-(1/\rho)\|\boldsymbol{y}^{k}-\boldsymbol{y}^{k+1}\|_2^2
\end{aligned}\tag{5-18}$$

Taking summation of above inequalities and noticing $L_\rho(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v},\boldsymbol{y})$ is bounded below, it gives

$$\begin{aligned}
&c\sum_{k=0}^{+\infty}\left(\|\boldsymbol{x}^{k}-\boldsymbol{x}^{k+1}\|_2^2+\|\boldsymbol{u}^{k}-\boldsymbol{u}^{k+1}\|_2^2+\|\boldsymbol{v}^{k}-\boldsymbol{v}^{k+1}\|_2^2\right)\\
&-(1/\rho)\sum_{k=0}^{+\infty}\|\boldsymbol{y}^{k}-\boldsymbol{y}^{k+1}\|_2^2<+\infty
\end{aligned}\tag{5-19}$$

According to assumption ②, we have $\sum_{k=0}^{+\infty}\|\boldsymbol{y}^{k}-\boldsymbol{y}^{k+1}\|_2^2<+\infty$. Therefore

$$\sum_{k=0}^{+\infty}\left(\|\boldsymbol{x}^{k}-\boldsymbol{x}^{k+1}\|_2^2+\|\boldsymbol{u}^{k}-\boldsymbol{u}^{k+1}\|_2^2+\|\boldsymbol{v}^{k}-\boldsymbol{v}^{k+1}\|_2^2\right)<+\infty\tag{5-20}$$

which implies the convergence of $\{(\boldsymbol{x}^{k},\boldsymbol{u}^{k},\boldsymbol{v}^{k})\}$. We then denote the limit point as $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}})$. In other words, $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})$ is a stationary point of ADMM-IPM-STO. Therefore, condition ①, ② and ③ in Theorem 5.1 are satisfied. Furthermore, assumption ① directly leads to condition ④ in Theorem 5.1. So $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}})$ is a local minimum of problem (5-5). ■

Remark 5.2： Theorem 5.2 actually reveals that if the convergence of the dual variable $\boldsymbol{y}^{k}$ is observed, it is safe to say that ADMM-IPM-STO achieves a local minimum of problem (5-5). This theorem substantiates the optimality of the solutions obtained by ADMM-IPM-STO in practice. □

Finally, since $g(\boldsymbol{v})$ is convex for $q=1$, we can obtain stronger convergence result under proper assumptions. We present the strong result in the following theorem.

Theorem 5.3： Let $\{(\boldsymbol{x}^{k},\boldsymbol{u}^{k},\boldsymbol{v}^{k},\boldsymbol{y}^{k})\}$ be a sequence generated by ADMM-IPM-STO. $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})$ satisfies the sufficient optimality in Theorem 5.1. Assume that, for sufficient large k: ① problem $min\ f(\boldsymbol{x},\boldsymbol{u})+\left(\boldsymbol{y}^{k}\right)^{T}\boldsymbol{u},\ s.t.\ (\boldsymbol{x},\boldsymbol{u})\in S$ is generic; ② $(\boldsymbol{x}^{k},\boldsymbol{u}^{k})$ is in the attraction basin of local minimum $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}})$ of $L_\rho(\boldsymbol{x},\boldsymbol{u},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})$, i.e. $L_\rho(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})\leq L_\rho(\boldsymbol{x}^{k},\boldsymbol{u}^{k},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})$ always holds; ③ $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}})$ is in the attraction basin of local minimum $(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1})$ of $L_\rho(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v}^{k},\boldsymbol{y}^{k})$, i.e. $L_\rho(\boldsymbol{x}^{k+1},\boldsymbol{u}^{k+1},\boldsymbol{v}^{k},\boldsymbol{y}^{k})\leq L_\rho(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\boldsymbol{v}^{k},\boldsymbol{y}^{k})$ always holds. Then $\{(\boldsymbol{x}^{k},\boldsymbol{u}^{k},\boldsymbol{v}^{k},\boldsymbol{y}^{k})\}$ converges to $(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}},\overline{\boldsymbol{v}},\overline{\boldsymbol{y}})$. □

Proof： Since $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}})$ is a local minimum of $L_\rho(\boldsymbol{x}, \boldsymbol{u}, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}})$ and $(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1})$ is in its attraction basin by assumption ②, we have $L_\rho(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}}) \leq L_\rho(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1}, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}})$. Considering assumption ① and using Lemma 1, we obtain

$$f(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}) + \overline{\boldsymbol{y}}^T \overline{\boldsymbol{u}} \leq f(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1}) + \overline{\boldsymbol{y}}^T \boldsymbol{u}^{k+1} \tag{5-21}$$

Similarly, $\overline{\boldsymbol{v}}$ is the global minimum of $L_\rho(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \boldsymbol{v}, \overline{\boldsymbol{y}})$, so we have $L_\rho(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}}) \leq L_\rho(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \boldsymbol{v}^{k+1}, \overline{\boldsymbol{y}})$. Considering the convexity of $g(\boldsymbol{v})$ and using Lemma 5.1, we obtain

$$g(\overline{\boldsymbol{v}}) - \overline{\boldsymbol{y}}^T \overline{\boldsymbol{v}} \leq g(\boldsymbol{v}^{k+1}) - \overline{\boldsymbol{y}}^T \boldsymbol{v}^{k+1} \tag{5-22}$$

Add (5-21) and (5-22) and notice $\overline{\boldsymbol{u}} - \overline{\boldsymbol{v}} = \boldsymbol{0}$ to obtain

$$f(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}) + g(\overline{\boldsymbol{v}}) \leq f(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1}) + g(\boldsymbol{v}^{k+1}) + \overline{\boldsymbol{y}}^T \left(\boldsymbol{u}^{k+1} - \boldsymbol{v}^{k+1}\right) \tag{5-23}$$

According to ADMM-IPM-STO, $(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1})$ is the local minimum of $L_\rho(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{v}^k, \boldsymbol{y}^k)$. Considering assumption ③, substituting $\boldsymbol{y}^k = \boldsymbol{y}^{k+1} - \rho(\boldsymbol{u}^{k+1} - \boldsymbol{v}^{k+1})$ and again using Lemma 5.1, we obtain

$$\begin{aligned} &f(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1}) + \left(\boldsymbol{y}^{k+1} + \rho(\boldsymbol{v}^{k+1} - \boldsymbol{v}^k)\right)^T \boldsymbol{u}^{k+1} \\ &\quad \leq f(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}) + \left(\boldsymbol{y}^{k+1} + \rho(\boldsymbol{v}^{k+1} - \boldsymbol{v}^k)\right)^T \overline{\boldsymbol{u}} \end{aligned} \tag{5-24}$$

Similarly, noticing $\boldsymbol{v}^{k+1}$ is the global minimum of $L_\rho(\boldsymbol{x}^{k+1}, \boldsymbol{u}^{k+1}, \boldsymbol{v}, \boldsymbol{y}^k)$, substituting $\boldsymbol{y}^k = \boldsymbol{y}^{k+1} - \rho(\boldsymbol{u}^{k+1} - \boldsymbol{v}^{k+1})$ and using Lemma 5.1 yields

$$g(\boldsymbol{v}^{k+1}) - \left(\boldsymbol{y}^{k+1}\right)^T \boldsymbol{v}^{k+1} \leq g(\overline{\boldsymbol{v}}) - \left(\boldsymbol{y}^{k+1}\right)^T \overline{\boldsymbol{v}} \tag{5-25}$$

In the same way, we also have

$$g(\boldsymbol{v}^{k+1}) - \left(\boldsymbol{y}^{k+1}\right)^T \boldsymbol{v}^{k+1} \leq g(\boldsymbol{v}^k) - \left(\boldsymbol{y}^{k+1}\right)^T \boldsymbol{v}^k \tag{5-26}$$

$$g(\boldsymbol{v}^k) - \left(\boldsymbol{y}^k\right)^T \boldsymbol{v}^k \leq g(\boldsymbol{v}^{k+1}) - \left(\boldsymbol{y}^k\right)^T \boldsymbol{v}^{k+1} \tag{5-27}$$

We add (5-26) and (5-27) to obtain

$$(\boldsymbol{y}^{k+1} - \boldsymbol{y}^k)^T (\boldsymbol{v}^{k+1} - \boldsymbol{v}^k) \geq 0 \tag{5-28}$$

Adding (5-23), (5-24) and (5-25), rearranging (see Appendix F), and noticing (5-28), we can obtain

$$w^k - w^{k+1} \geq \rho \|\boldsymbol{u}^{k+1} - \boldsymbol{v}^{k+1}\|_2^2 + \rho \|\boldsymbol{v}^{k+1} - \boldsymbol{v}^k\|_2^2 \tag{5-29}$$

where $w^k = (1/\rho)\|\boldsymbol{y}^k - \overline{\boldsymbol{y}}\|_2^2 + \rho \|\boldsymbol{v}^{k+1} - \boldsymbol{v}^k\|_2^2$. This shows that w^k monotonously decreases in each iteration until the residuals vanish. Therefore $\{(\boldsymbol{x}^k, \boldsymbol{u}^k, \boldsymbol{v}^k, \boldsymbol{y}^k)\} \to (\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}, \overline{\boldsymbol{v}}, \overline{\boldsymbol{y}})$.

Adding (5-23), (5-24) and (5-25) and multiplying by 2, we obtain

$$2\left(\boldsymbol{y}^{k+1}-\overline{\boldsymbol{y}}\right)^{T}\boldsymbol{r}^{k+1}+2\rho\left(\boldsymbol{v}^{k+1}-\boldsymbol{v}^{k}\right)^{T}\left(\boldsymbol{u}^{k+1}-\overline{\boldsymbol{u}}\right)\leq 0$$

where $\boldsymbol{r}^{k+1}=\boldsymbol{u}^{k+1}-\boldsymbol{v}^{k+1}$. By noticing $\boldsymbol{y}^{k+1}=\boldsymbol{y}^{k}+\rho\boldsymbol{r}^{k+1}$ and $\overline{\boldsymbol{u}}-\overline{\boldsymbol{v}}=\mathbf{0}$, we rewrite the left hand side (LHS) of the above inequality as follows:

$$\begin{aligned}\text{LHS}=&2\left(\boldsymbol{y}^{k}+\rho\boldsymbol{r}^{k+1}-\overline{\boldsymbol{y}}\right)^{T}\boldsymbol{r}^{k+1}+2\rho\left(\boldsymbol{v}^{k+1}-\boldsymbol{v}^{k}\right)^{T}\left(\boldsymbol{r}^{k+1}\right.\\&\left.+\boldsymbol{v}^{k+1}-\overline{\boldsymbol{v}}\right)\\=&\left\{2\left(\boldsymbol{y}^{k}-\overline{\boldsymbol{y}}\right)^{T}\boldsymbol{r}^{k+1}+\rho\|\boldsymbol{r}^{k+1}\|_{2}^{2}\right\}+\left\{\rho\|\boldsymbol{r}^{k+1}\|_{2}^{2}+\right.\\&2\rho\left(\boldsymbol{v}^{k+1}-\boldsymbol{v}^{k}\right)^{T}\boldsymbol{r}^{k+1}+2\rho\left(\boldsymbol{v}^{k+1}-\boldsymbol{v}^{k}\right)^{T}\left(\boldsymbol{v}^{k+1}-\boldsymbol{v}^{k}\right)\\&\left.+2\rho\left(\boldsymbol{v}^{k+1}-\boldsymbol{v}^{k}\right)^{T}\left(\boldsymbol{v}^{k+1}-\overline{\boldsymbol{v}}\right)\right\}\\=&\frac{1}{\rho}\left(\|\boldsymbol{y}^{k+1}-\overline{\boldsymbol{y}}\|_{2}^{2}-\|\boldsymbol{y}^{k}-\overline{\boldsymbol{y}}\|_{2}^{2}\right)+\rho\left(\|\boldsymbol{v}^{k+1}-\overline{\boldsymbol{v}}\|_{2}^{2}\right.\\&\left.-\|\boldsymbol{v}^{k}-\overline{\boldsymbol{v}}\|_{2}^{2}\right)+\rho\|\boldsymbol{r}^{k+1}+(\boldsymbol{v}^{k+1}-\boldsymbol{v}^{k})\|_{2}^{2}\end{aligned}$$

Noticing the positivity of $(\boldsymbol{r}^{k+1})^{T}(\boldsymbol{v}^{k+1}-\boldsymbol{v}^{k})$ according to (5-28), LHS ≤ 0 directly leads to (5-29). ■

Remark 5.3： Theoretically, the IPM algorithm for OPF problems can only achieve a local minimum. So assumption ② and ③ in theorem 5.3 actually ensure that all the local minimums are in a single attraction basin and thus the inequality relations do not compromise. □

5.3 General FACTS Devices Allocation Problems

In our formulation, current mismatch equations are chosen as equality constraints instead of power mismatch equations. Bus voltages and generator current injections are taken as state variables. Consequently, generators and loads are modeled as complex current injections at their buses. All FACTS devices are modeled as parametric complex current injections at related buses. The reasons for these choices are as follows. At first, three series controllable parameters relate to one line in general FACTS devices allocation problems, which leads to a very high-order power balance equation. Subsequently, solving the second-order derivatives is far more difficult in power balance equations than that in current balance equations. Secondly, as every bus or line can be a candidate location for FACTS devices implies a very large number of controllable parameters, and solving the second-order derivatives in conventional formulation thus becomes rather impractical.

5.3.1 Branch and FACTS Devices Modeling

Without loss of generality, every bus or line in power systems is considered as a candidate location for FACTS device placement. A general branch model is shown in Fig.1 which is sim-

ilar to that in [116]. r, x and b are transmission line parameters. $\hat{v}_f$, $\hat{v}_t$, $\hat{i}_f$ and $\hat{i}_t$ are complex voltage and current at "from" and "to" ends of the branch. The series controllable parameters κ, φ and τ are used to describe the effect of TCSC (thyristor controlled series capacitor), TCPS (thyristor controlled phase shifter) and ULTC (under load tap changer), respectively. In addition, Shunt compensation devices can be simply modeled as extra susceptance at certain buses which are not shown in Fig. 1. Because this model contains all network parameters, other types of FACTS devices can be equivalently transformed into this model. In particular, the models of STATCOM [150], SSSC [151] and UPFC [152] are special cases of the general model used in this chapter.

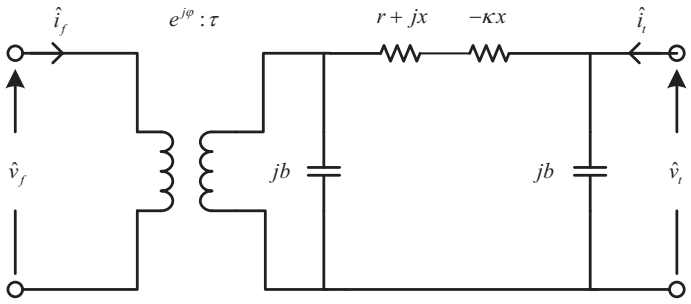


Figure 5-1 General Branch Model

From Fig. 1, the relationship between complex voltages and currents at both ends is given by

$$\begin{bmatrix}\hat{i}_f\\ \hat{i}_t\end{bmatrix}=\begin{bmatrix}\tau^2(\hat{y}^*+jb) & -\tau e^{j\varphi}\hat{y}^*\\ -\tau e^{-j\varphi}\hat{y}^* & \hat{y}^*+jb\end{bmatrix}\begin{bmatrix}\hat{v}_f\\ \hat{v}_t\end{bmatrix} \tag{5-30}$$

where $\hat{y}^*=1/\left(r+j(1-\kappa)x\right)$.

To facilitate optimization computation, the equivalent current injection model is derived and adopted. In this model, the effects of all the controllable parameters are represented by the current injections at the "from" and "to" ends of the branch and this model possesses the versatility to adapt to various types of FACTS devices. Assume that Fig. 1 is equivalent to Fig. 2 in which equivalent complex current injections $\Delta\hat{i}_f$ and $\Delta\hat{i}_t$ are introduced to depict the effects of all FACTS devices and $\hat{y}=1/\left(r+jx\right)$. The following equation holds:

$$\begin{bmatrix}\hat{i}_f+\Delta\hat{i}_f\\ \hat{i}_t+\Delta\hat{i}_t\end{bmatrix}=\begin{bmatrix}\hat{y}+jb & -\hat{y}\\ -\hat{y} & \hat{y}+jb\end{bmatrix}\begin{bmatrix}\hat{v}_f\\ \hat{v}_t\end{bmatrix} \tag{5-31}$$

Combining (5-30) with (5-31), we can acquire parametric complex current injections as

$$\begin{bmatrix}\Delta\hat{i}_f\\ \Delta\hat{i}_t\end{bmatrix}=\begin{bmatrix}\hat{y}-\tau^2\hat{y}^*+jb(1-\tau^2) & -\hat{y}+\tau e^{j\varphi}\hat{y}^*\\ -\hat{y}+\tau e^{-j\varphi}\hat{y}^* & \hat{y}-\hat{y}^*\end{bmatrix}\begin{bmatrix}\hat{v}_f\\ \hat{v}_t\end{bmatrix} \tag{5-32}$$

Equation (5-32) shows that $\Delta\hat{i}_f$ and $\Delta\hat{i}_t$ are the functions of series controllable parameters

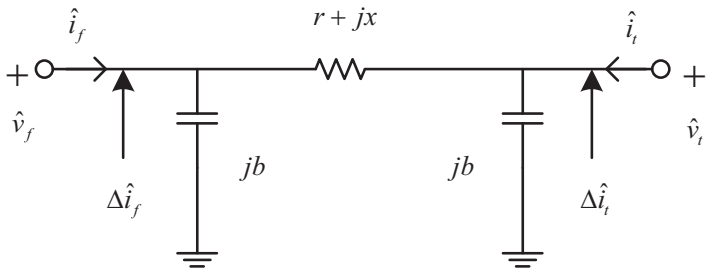


Figure 5-2 Equivalent Current Injection Model

and complex voltages at both ends of the line and they can completely represent the effect of controllable devices.

5.3.2 Sparsity-constrained OPF Model

In the sparsity-constrained OPF problems, decision variables are divided into two groups and denoted by two vectors. Decision vector $\boldsymbol{u}$ consists of setting values of all the candidate FACTS devices, i.e.

$$\boldsymbol{u} = (\boldsymbol{b}_{sh},\ \boldsymbol{\varphi},\ \boldsymbol{\kappa}) \tag{5-33}$$

where $\boldsymbol{b}_{sh} \in \mathbf{R}^{n_b}$ denotes the susceptance of shunt compensation devices at every bus; $\boldsymbol{\varphi} \in \mathbf{R}^{n_l}$ and $\boldsymbol{\kappa} \in \mathbf{R}^{n_l}$ denote the shift angle and series compensation rate at every line, respectively. Certain component of $\boldsymbol{u}$ being zero indicates no corresponding device is installed. Vector $\boldsymbol{x}$ contains bus voltages, generator current injections and UTLC ratios, i.e.

$$\boldsymbol{x} = (\boldsymbol{e},\ \boldsymbol{f},\ \boldsymbol{a}_g,\ \boldsymbol{b}_g,\ \boldsymbol{\tau}) \tag{5-34}$$

where $\boldsymbol{e} \in \mathbf{R}^{n_b}$ and $\boldsymbol{f} \in \mathbf{R}^{n_b}$ denote the real and imaginary parts of bus voltages; $\boldsymbol{a}_g \in \mathbf{R}^{n_g}$ and $\boldsymbol{b}_g \in \mathbf{R}^{n_g}$ denote the real and imaginary parts of generator current injections; $\boldsymbol{\tau} \in \mathbf{R}^{n_l}$ denotes the transformer ratio at every line. Between them, vector $\boldsymbol{u}$ is expected to be very sparse after optimization due to economic consideration.

The equality constraints of OPF formulation in this chapter are the network current mismatch equations. Let $\hat{\boldsymbol{h}} \in \mathbf{C}^{n_b}$ denotes the complex current mismatch at every bus. Its i-th element is given by

$$\hat{h}_i(\boldsymbol{x}, \boldsymbol{u}) = \sum_{k \in \mathcal{G}_i} (a_{gk} + jb_{gk}) + \left(\frac{p_{li} + jq_{li}}{e_i + jf_i}\right)^c + \Delta\hat{i}_i - \sum_{k=1}^{n_b} \hat{y}_{ik}(e_k + jf_k) \tag{5-35}$$

where $\mathcal{G}_i$ is the index set of generators installed at bus i; p_{li} and q_{li} denote the active and reactive power of load at bus i, respectively; $\hat{y}_{ik}$ is the (i,k)-th element of the nodal admittance matrix of the original network shown in Fig. 2. $\Delta\hat{i}_i$ is the complex current injection induced by all the controllable equipments related to bus i. Define $\mathcal{F}_i$ $(\mathcal{T}_i)$ as the index set of branches which

takes bus i as their "from" ("to") end. Then $\Delta\hat{i}_i$ is expressed as

$$\Delta\hat{i}_i = \sum_{k\in\mathcal{F}_i} \Delta\hat{i}_{fk} + \sum_{k\in\mathcal{T}_i} \Delta\hat{i}_{tk} - jb_{shi}(e_i + jf_i) \tag{5-36}$$

where $\Delta\hat{i}_{fk}$ ($\Delta\hat{i}_{tk}$) is the complex current injection induced by related controllable devices at the "from" ("to") end of the branch k which is given by (5-32). The last term in (5-36) denotes the complex current injection induced by shunt compensation devices at bus i. Therefore, the equality constraints can be written in a compact form as

$$\boldsymbol{h}(\boldsymbol{x},\boldsymbol{u}) = \left(\text{Re}\left(\hat{\boldsymbol{h}}(\boldsymbol{x},\boldsymbol{u})\right),\ \text{Im}\left(\hat{\boldsymbol{h}}(\boldsymbol{x},\boldsymbol{u})\right)\right) = \boldsymbol{0} \tag{5-37}$$

The inequality constraints are steady state security constraints, including current magnitude limit for every line and voltage magnitude limit for every bus

$$\|(\hat{y}+jb)(e_f+jf_f) - \hat{y}(e_t+jf_t)\|_2^2 \le (i_{max})^2 \tag{5-38}$$

$$(v_{mim})^2 \le \|e_i + jf_i\|_2^2 \le (v_{max})^2 \tag{5-39}$$

and physical limits of devices, including generator active and reactive power output:

$$p_{kmim} \le \text{Re}\left((e_i+jf_i)(a_{gk}+jb_{gk})^c\right) \le p_{kmax} \tag{5-40}$$

$$q_{kmim} \le \text{Im}\left((e_i+jf_i)(a_{gk}+jb_{gk})^c\right) \le q_{kmax} \tag{5-41}$$

where generator k is installed at bus i; setting value limits of ULTC, SVC, TCPS and TCSC:

$$\tau_{kmim} \le \tau_k \le \tau_{kmax} \tag{5-42}$$

$$b_{shkmim} \le b_{shk} \le b_{shkmax} \tag{5-43}$$

$$\varphi_{kmim} \le \varphi_k \le \varphi_{kmax} \tag{5-44}$$

$$\kappa_{kmim} \le \kappa_k \le \kappa_{kmax} \tag{5-45}$$

Thus inequality constraints are written in a compact form as

$$\boldsymbol{g}_{mim} \le \boldsymbol{g}(\boldsymbol{x},\boldsymbol{u}) \le \boldsymbol{g}_{max} \tag{5-46}$$

In the proposed formulation, it is certainly free to choose various kinds of objective functions. Without loss of generality, system loadability is taken as objective function in our case studies to test the proposed algorithm. The complex power injection of loads at bus i is modified as

$$p_{li} + jq_{li} = \eta(p_{li0} + jq_{li0}) \tag{5-47}$$

where η is the loadability factor and $p_{li0} + jq_{li0}$ is the initial complex power injection of loads at bus i. Hence, the objective function is given by

$$f(\boldsymbol{x}, \boldsymbol{u}) = -\eta \tag{5-48}$$

To sum up, the sparsity-constrained OPF formulation for general FACTS devices allocation problems can be written in a compact form as (5-3) where feasible set S are specified by (5-37) and (5-46). Objective function is defined in (5-48). Problem (5-3) can be solved with the ADMM-IPM-STO algorithm discussed in Section 5.3.

5.4 Case Studies

To validate the algorithm proposed, single-type and multiple-type FACTS devices allocation is tested on standard IEEE 30-, 118- and 300-bus systems, respectively. The system data is extracted from Matpower 4.1. ADMM-IPM-STO algorithm was programmed in MATLAB running on a Win7 PC with Intel Core i5 1.80-GHz CPU and 4 GB of RAM.

5.4.1 Parameter Settings

The value of q of L_q norms is an economical parameter related to the investment cost of different types of FACTS device. For practical application, L_q norms can choose three values, $L_{1/2}$, $L_{2/3}$ and L_1. If the investment cost of a certain type of FACTS devices is dominated by the number of devices, $L_{1/2}$ norm is a better choice. If the setting value of the FACTS devices play a major role in the investment cost, L_1 norm is preferred. $L_{2/3}$ norm acts as a compromise between $L_{1/2}$ norm and L_1 norm. In our experiment, $L_{1/2}$ norm is used in multiple-type FACTS devices allocation and single-type TCPS allocation; L_1 norm is applied to single-type SVC allocation and TCSC allocation.

The physical meaning of regularization parameter λ represents the cost of per unit FACTS devices measured in L_q norm and its value should be decided by the decision makers based on their specification. To test the robustness of the algorithm proposed, the sparsity-constrained OPF problems are solved with decreasing values of λ, which means system loadability is increased through installation of an increasing number of FACTS devices. In other words, every λ value is associated with a devices allocation strategy. Due to space limitation, only under a certain device number, the allocation strategy with the largest loadability factor is given to illustrate the relationship among device numbers, total installed capacity and system loadability.

Theoretically, the value of the augmented Lagrangian parameter ρ will not affect the result of the algorithm only if it surpasses a threshold value which is problem-dependent and unknown before the problem solved. Nevertheless, the value of ρ affects the convergence process of the algorithm. Large values of ρ place a severe penalty on violations of primal feasibility and thus

tend to produce small primal errors. Conversely, small values of ρ tend to reduce the dual errors at the cost of the primal errors. Therefore, an ideal value of ρ should keep the primal and the dual errors within moderate difference as they both converge to zero. ρ is usually chosen by cross-validation.

In addition, the decision variable $\boldsymbol{u}$ is usually rescaled by multiplying its components with some factors. For example, considering different line distances, every κ is multiplied by the reactance of the related line. In multi-type FACTS devices allocation problems, to reflect the price differences among different types of FACTS devices, b_{sh}, κ and φ are multiplied by different factors α_c, α_κ and α_φ, respectively. In our study, the allowable range of the compensation rate of TCSCs, the shift angle of TCPSs and the ratio of UTLCs are $0\% \sim 50\%$, $-15^\circ \sim +15^\circ$ and $0.9 \sim 1.1$, respectively. The setting values of SVC are unbounded. ϵ_{primal} and ϵ_{dual} are both set to be 10^{-4}.

5.4.2 Multiple-type FACTS Devices Allocation

Multiple-type FACTS devices allocation is conducted on IEEE 30-bus system. The candidate FACTS devices are SVC, TCSC and TCPS. In this problem, $\boldsymbol{x}$ is 72 dimensional and $\boldsymbol{u}$ is 112 dimensional. By running a conventional OPF without FACTS devices, the maximum loadability of IEEE 30-bus system is 1.020. Regardless of device costs, i.e. setting $\lambda = 0$ in our algorithm, the theoretical maximum loadability by FACTS device installation is 1.735, and the results under this condition are taken as the initial values of decision variables in problems with other values of λ. Thus, during the optimization process, the loadability factor η, the device number, setting values decrease from the original values to achieve primal and dual convergence of the ADMM-IPM-STO algorithm. The convergence processes of primal and dual errors are shown in Fig. 3, and the changing process of the loadability factor is shown in Fig. 4 with $\lambda = 0.29$, $\rho = 500$, $\alpha_c = 0.1$, $\alpha_\kappa = 20$ and $\alpha_\varphi = 200$.

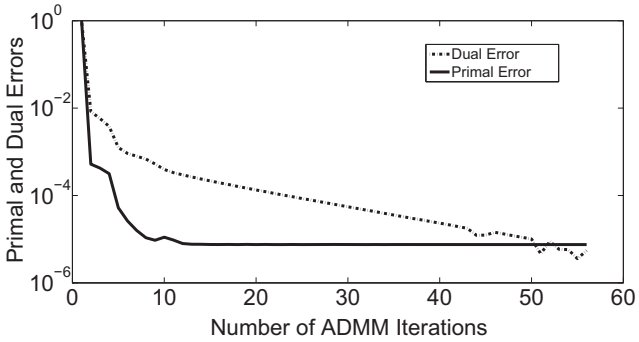


Figure 5-3 Primal and Dual Residuals Convergence Process

Fig.3 shows that the primal error stops decreasing after about 10 iterations, as the NLP sub-problem precision restriction is set to be 10^{-5}. The dual convergence is much slower than the primal convergence. Essentially, the reason for this phenomenon is that, in the method

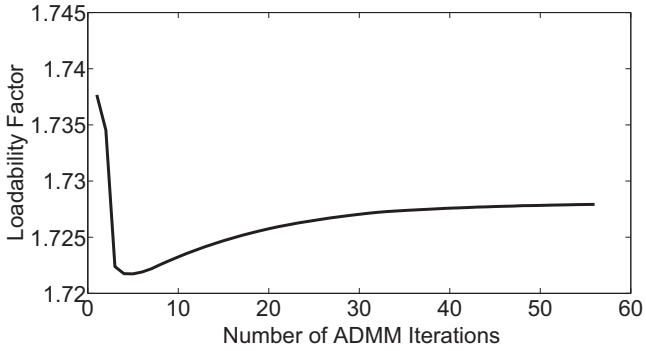


Figure 5-4 Loadability Factor Changing Process

of multipliers, the primal problem is solved with a second-order method (IPM) whereas the dual problem with a first-order method (steepest ascent) [153]. Fig. 4 shows the loadability factor declines from the initial value and finally reaches the steady value 1.728 as convergence achieved. After optimization, only two SVCs and two TCSCs, among 112 candidates, are selected and installed in the network, shown in the last row of Table I.

Furthermore, the allocation strategies with device number from 1 to 4 are listed in Table I. It is obvious that, with only 4 FACTS devices installed, the loadability factor can reach 1.728, up to 99.6% of the theoretical maximum. This validates the sparse feature of FACTS devices allocation problem. In other words, a large number of FACTS devices are not only uneconomical but also unnecessary.

Table 5-1 Multiple-Type FACTS Devices Allocation Strategy on 30-bus System

λ	No.	η	Allocation Strategy
3.38	1	1.539	TCSC:10(50%)①
2.67	2	1.608	SVC:8(90.5)②; TCSC:10(50%)
0.30	3	1.723	SVC:8(33.1); TCSC:10(36.0%),29(49.5%)
0.29	4	1.728	SVC:8(46.0),28(25.2); TCSC:10(34.1%),29(50%)

① Line number (compensation rate); ② Bus number (Var compensation capacity in MVar)

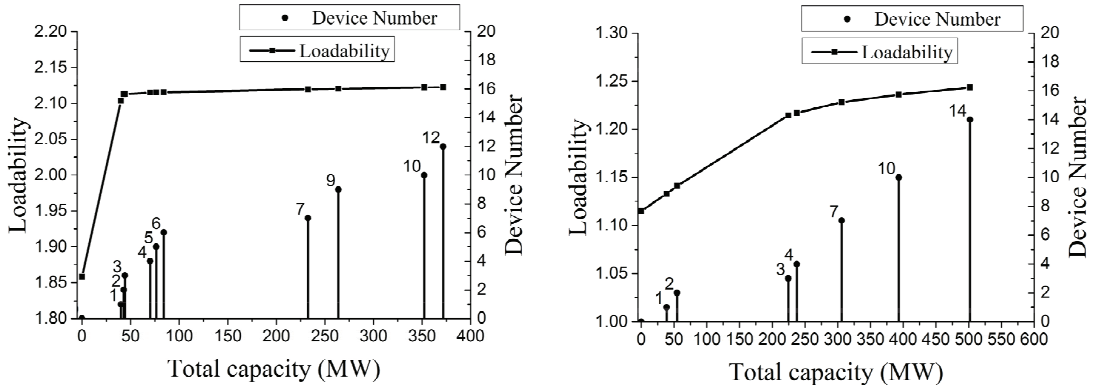
5.4.3 Single-Type FACTS Devices Allocation

SVC, TCSC and TCPS allocation problems are conducted on IEEE 118- and 300-bus systems. The basic information of all the six problems is summarized in Table II.

Results of all the six problems are shown in Fig.5 to Fig. 7, which illustrate the evolution of system loadability as the device number and total installed capacity increase. In all three types of FACTS allocation problems studied in this chapter, the maximum loadability saturates after several devices installed (typical number of devices is 3), and the marginal utility of FACTS devices significantly diminishes.

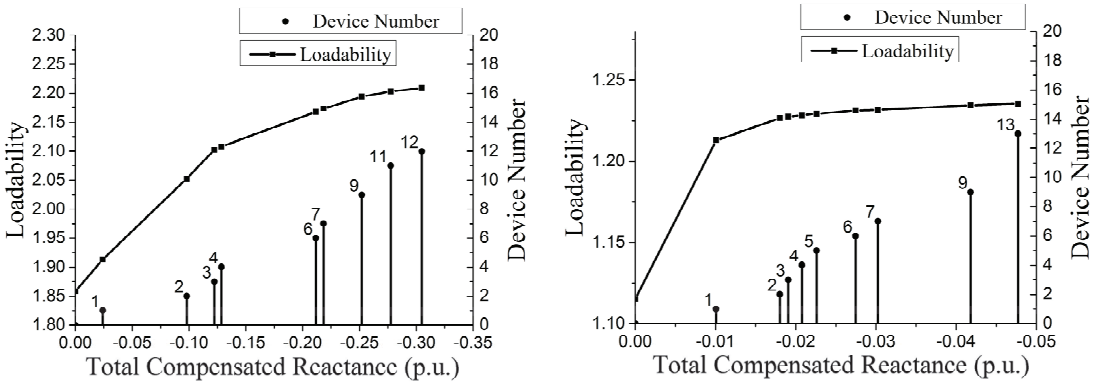
Table 5-2 Single-Type FACTS Devices Allocation Problems

Problems	Dim. $\boldsymbol{x}$	Dim. $\boldsymbol{u}$	η_{max} no device	theoretical η_{max}
SVC118	353	118	1.859	2.124
TCSC118	353	186	1.859	2.283
TCPS118	353	186	1.859	1.867
SVC300	845	300	1.116	1.243
TCSC300	845	411	1.116	1.236
TCPS300	845	411	1.116	1.120



(a) 118-bus system (b) 300-bus system

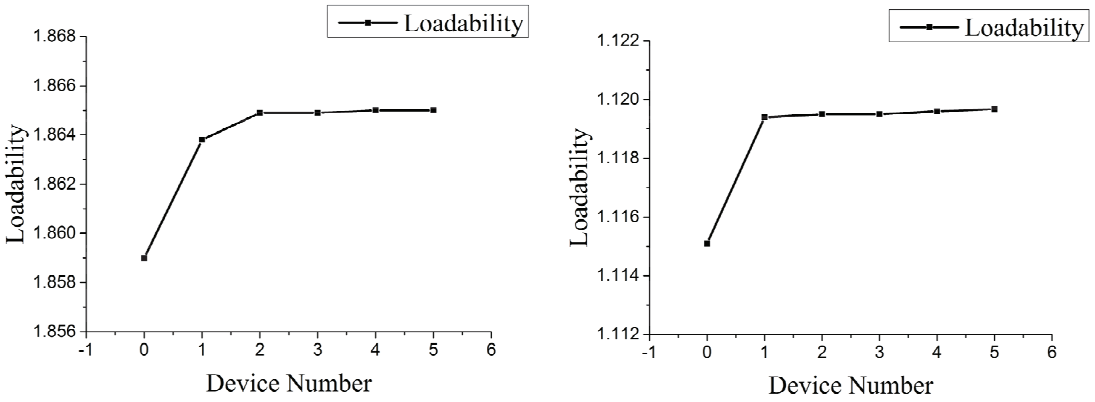
Figure 5-5 Loadability enhancement by optimal allocation of SVCs on 118 and 300 bus systems



(a) 118-bus system (b) 300-bus system

Figure 5-6 Loadability enhancement by optimal allocation of TCSCs on 118 and 300 bus systems

Detailed allocation strategies for SVC allocation on IEEE 118-bus system, TCSC allocation on IEEE 300-bus system and TCPS allocation on IEEE 118-bus system are listed in Table III, Table IV and Table V, respectively. Only the parameters and locations for the first 3 devices are given, as it is clearly shown in Fig.5 to Fig. 7 that the enhancement of loadability will saturate with more than 3 devices. This again demonstrates the sparsity of FACTS devices



(a) 118-bus system (b) 300-bus system

Figure 5-7 Loadability enhancement by optimal allocation of TCPSs on 118 and 300 bus systems

Table 5-3 SVC Allocation Strategy on 118-Bus System

No.	η	Allocation Strategy	η/η_{max}
1	2.103	76(40.12)③	99.0%
2	2.113	76(33.96), 118(8.89)	99.5%
3	2.114	76(34.06), 95(1.28), 118(8.82)	99.5%

③ Bus number (Var compensation capacity in MVar)

allocation problems.

Table 5-4 TCSC Allocation Strategy on 300-Bus System

No.	η	Allocation Strategy	η/η_{max}
1	1.213	177(37%)④	98.1%
2	1.226	177(50%), 367(26%)	99.2%
3	1.227	1(12%), 177(50%), 367(31%)	99.3%

④ Line number (compensation rate)

Table 5-5 TCPS Allocation Strategy on 118-Bus System

No.	η	Allocation Strategy	η/η_{max}
1	1.864	123(-5.418)⑤	99.8%
2	1.865	121(3.448), 123(-4.189)	99.9%
3	1.865	121(1.932), 122(1.590), 123(-4.113)	99.9%

⑤ Line number (shift angle in degree)

5.5 Result Discussion

5.5.1 Comparisons with Other Methods

Full AC power flow model has been used in our problem formulation. This certainly improves the quality of the solution compared to other methods employing simplified power flow models. The DC flow used in MILP method [132] neglects the nonlinearity of the power flow and constraints on voltage magnitude and cannot precisely reflect the limits on line capacity. The LFB flow used in MIQP method [122] has also been simplified by transforming the quadratic equality constraints into linear inequality constraints. Therefore the nonlinearity of power flow is not completely represented. The modeling of limits on line capacity is also deficient. The GA method in [128] uses a simplified version of AC flow by neglecting transverse conductance of transmission lines. The motivation for all those simplifications is either convenience of adopting certain methods or simplifying numerical computation. But this will inevitably affect the quality of the solutions. For example, we have observed in our experiment, the constraints of voltage magnitude and line capacity often act as binding constraints at the final solution which indicates that defects in representing these constraints surely change the optimal solution. These simplified methods are, to some extent, eligible for preliminary planning, but their results need to be verified by the full AC model. Although the different problem formulations have complicated influence on the final solutions, we would like to make some rough comparison as follows. For SVC allocation on 300-bus system, we install 3 SVCs to improve the loadability to 1.217 while [126] improves the loadability to 1.207 with 5 SVCs. For TCSC allocation on 300-bus system, the loadability is improved to 1.227 with 3 TCSCs in this chapter compared to 1.081 with 19 TCSCs in [122]. For TCPS allocation on 118-bus system, the loadability achieves 1.865 with 3 TCPSs by our approach compared to 1.76 with 13 TCPSs in [132]. These comparisons show that our approach is generally more effective to identify the optimal locations and setting values of FACTS devices.

To roughly evaluate the efficiency of the proposed algorithm, we continue the comparison originally conducted in [122] and list the results in Table VI. At first, the proposed ADMM-IPM-STO employs the full AC power flow model rather than simplified models such as the DC model and the LFB model used in MILP [132] and MIQP [122], which improves the accuracy and reliability of the computation results. Secondly, the ADMM-IPM-STO is far more computationally efficient than GA [128] and also offers faster or at least comparable performance compared with MIQP [122].

The major competitors of the proposed method are those mixed-integer programming based methods, including MILP [132], MIQP [122] and MINP [126]. In those methods, branch-and-cut or benders decomposition are involved to tackle the binary variables and form

Table 5-6 Algorithm Efficiency Comparison

Methods	Network	Device	Model	Machine	Time
GA[128]	200-bus	TCPS	AC	Sun SPARC Workstation	1.5h
MILP[132]	300-bus	TCPS	DC	PC 450MHz 128MB RAM	2.5s
MIQP[122]	300-bus	TCSC	LFB	Dell OptiPlex GX520	218s
Proposed	300-bus	TCPS	AC	PC 1.80GHz 3.85GB RAM	63s

a series of continuous subproblems. The continuous subproblems can be linear programming, quadratic programming or nonlinear programming according to their problem formulations. In those methods, the number of continuous subproblems is strongly related to the number of binary variables. Thus the computation time will significantly increase as the allowable device number or the system scale increase. We compare the computation time of the proposed method with MINP and MIQP in Fig.8 and Fig.9 as allowable device number and system scale increase, respectively. These two graphs show that the computation time of mixed-integer programming based methods are very sensitive to allowable device number and system scale. Whereas the computation time of ADMM-IPM-STO is almost irrelevant to allowable device number and far less sensitive to problem scale.

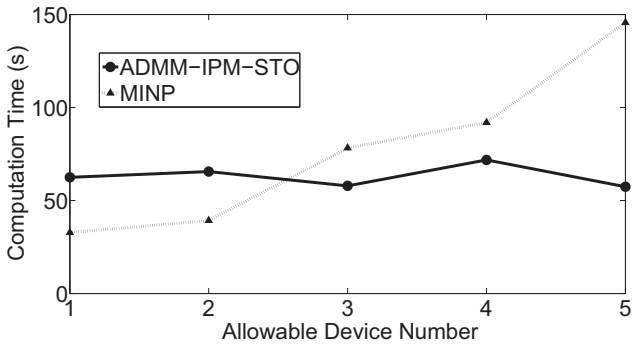


Figure 5-8 Computation time comparison between ADMM-IPM-STO and MINP as allowable device number increase for SVC allocation on 300-bus system

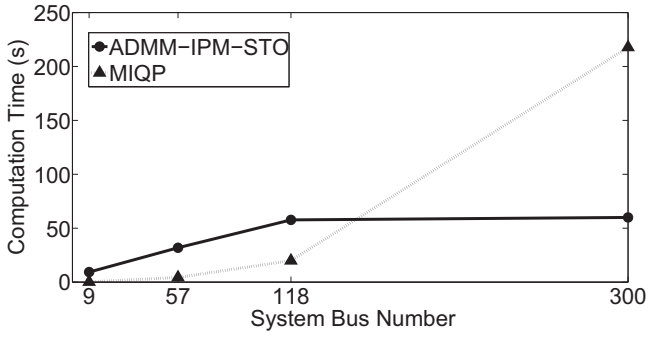


Figure 5-9 Computation time comparison between ADMM-IPM-STO and MIQP as system scale increase for TCSC allocation problem

5.5.2 Convergence and Non-convexity

Although so far general convergence result of ADMM-IPM-STO for all $q \in (0, 1]$ has not been established, our analysis shows the optimality of solutions obtained in practice is theoretically substantiated, and the convergence for $q = 1$ is quite well guaranteed. Furthermore, the convergence of proposed algorithm is very robust in practice for each value of q with closed-form solutions, including $q = 1/2$, $q = 2/3$ and $q = 1$. In our experiment, ADMM-IPM-STO can achieve convergence for every type of FACTS allocation problems as long as the parameter ρ is adjusted to a proper range. The rule of thumb of finding the proper ρ, in our experience, is to make the primal convergence a little faster than dual convergence. Actually, by only several trial-and-error processes a suitable parameter can then be found because the range of applicable ρ is usually very wide and the dependency of the algorithm performance on the value of ρ is pretty loose.

Theoretically, ADMM-IPM-STO can only achieve local minimum due to the non-convex problem formulation. Note that there are two sources of non-convexity: the original non-convexity of OPF problems and the non-convexity of L_q norms. For the non-convexity of OPF problems, because we have applied IPM to solve the OPF subproblems, ADMM-IPM-STO shares the same limitation of the IPM, and therefore only local minimum can be expected. Recently, there have been some attempts to obtain global solution of OPF problems. The author of [154] established a sufficient condition for zero duality gap of the semi-definite programming formulation of the dual problem of OPF, which leads to global solution to OPF problems. But this sufficient condition does not always hold in every network [155]. None of the current methods can guarantee the global optimum solution of general OPF problems. The proposed ADMM-IPM-STO thus is also incapable of guaranteeing global solution. For the additional non-convexity of L_q norms, a natural question is that whether this additional non-convexity will produce additional local optimums. We have conducted a series of numerical experiments to answer this question. For all tests, we fix the initial value of $\boldsymbol{x}$ and randomly select different initial values of $\boldsymbol{u}$. This means the initial condition of the original network is fixed while the initial guess of the FACTS device allocation strategy changes. Test results show that all different initial values of $\boldsymbol{u}$ result in a common optimal solution. This shows that the non-convexity brought by L_q does not produce additional local minimum. The fundamental reason for this is that the STOs give global optimal solutions to L_q regularization problems.

5.6 Conclusion

This chapter has proposed a novel formulation and algorithm for FACTS devices allocation problems. Based on the sparse characteristics of device placement, FACTS allocation problems have been formulated as a sparsity-constrained OPF problem. An ADMM-IPM-

STO algorithm, which combines the state-of-art algorithms in both sparse optimization and OPF, has been proposed to simultaneously determine the numbers, locations, setting values and types of FACTS devices. $L_q(0 < q < 1)$ norms have been firstly introduced to represent sparsity in FACTS devices allocation problems, in which q is an economical parameter related to the investment cost of different types of FACTS device. The proposed method has been tested on several IEEE standard systems in both multiple-type allocation problem as well as single-type problems, respectively. Case studies demonstrate the effectiveness and efficiency of this method.

In addition, the proposed formulation and algorithm can be easily applied to other objective functions since we have presented a general sparsity-constrained OPF model and the same calculating process of the ADMM-IPM-STO can be used no matter what objective function is chosen. Moreover, the proposed approach can be extended to FACTS allocation problems considering multiple contingencies by further incorporating the ideas of "security-constrained OPF" [149] and "group sparsity" [156]. FACTS devices allocation problem considering multiple contingencies can be formulated as

$$\min_{(\boldsymbol{x}_c,\boldsymbol{u}_c):\forall c\in\mathcal{C}} f(\boldsymbol{x}_0,\boldsymbol{u}_0)+\lambda\sum_{k=1}^{n}\|\boldsymbol{w}_k\|_p^q \tag{5-49a}$$

$$\boldsymbol{h}_c(\boldsymbol{x}_c,\boldsymbol{u}_c)=\boldsymbol{0} \qquad \forall c\in\mathcal{C} \tag{5-49b}$$

$$\boldsymbol{g}_{mim}\le\boldsymbol{g}_c(\boldsymbol{x}_c,\boldsymbol{u}_c)\le\boldsymbol{g}_{max} \qquad \forall c\in\mathcal{C} \tag{5-49c}$$

where $\mathcal{C}=\{0,1,2,\ldots,t\}$ is the set of prespecified $t+1$ contingencies with $c=0$ representing the base case; for each contingency, $\boldsymbol{u}_c$, $\boldsymbol{x}_c$, $\boldsymbol{h}_c(\cdot)$ and $\boldsymbol{g}_c(\cdot)$ share the same definition as (5-33), (5-34), (5-37) and (5-46), respectively; $\boldsymbol{w}_k=(u_{0k},u_{1k},\ldots,u_{tk})$ are the setting values of the k-th device in all contingencies. The regularization term in the objective function is referred as $L_{q,p}$ norm with $q\in(0,1]$ and $p\in[2,+\infty]$. This regularizer leads the setting values of the same FACTS device in different contingencies to be zeros or non-zeros simultaneously which enforces sparsity on installation sites.

6 Moment-SOS Approach to Interval Power Flow

6.1 Introduction

Integration of intermittent renewable sources and advent of deregulated competitive power markets increase the uncertainties of power system operation. It will be more and more difficult to acquire reliable information of steady-state operation by just running the conventional power flow (PF) program. There is a long-recognized and pressing need for power flow analysis tools to consider uncertainties.

Power flow analysis methods in the presence of uncertainty can be classified into two groups according to how uncertainty is represented. The first group is the probabilistic power flow (PPF) in which loads and generations are expressed as random variables with associated distribution functions. PPF aims at deriving the probability distribution of power flow solutions. Monte Carlo simulation based techniques [5, 6], analytical methods [7–9] and point estimate methods [10, 11] are extensively investigated to deal with this problem. However, this group of methods depends on the assumption that the random behavior of uncertainty obeys pre-defined distribution which is difficult to identify in practice.

The second group of methods, interval power flow (IPF) [16–20], model uncertainty of loads and generations as intervals without distribution structures, which seems a more practical approach because system operators need to consider the worst-case scenario to guarantee the security of the power networks. Interval arithmetic (IA) is introduced in [16] to modify conventional Newton iterations to obtain outer approximation of the solution set, but the results are highly conservative and the convergence is unfounded. A local search procedure is proposed in [17] to find accurate boundary solutions but the convergence is not proved and only local optimality is guaranteed. To overcome the limitation of the IA method, affine arithmetic (AA) is used in [18] by expressing power system state variables as affine function of uncertainties, and linear programming (LP) is employed to obtain tight bounds of uncertainty parameters. However, this method is still approximate in nature. In [19], from a range arithmetic perspective, a nonlinear programming (NLP) model with complementarity constraints is proposed to find the upper and lower bounds of power flow solutions. Although the formulation is non-convex and difficult to solve, this work provides a theory foundation for optimization-based method to solve IPF. In the most recent work [20], IPF is formulated as a quadratically constrained quadratic programming (QCQP) problem. Convex envelops are employed to relax the non-convex QCQP to a convex LP. Starting from estimated solution intervals, the optimality-based bounds tightening (OBBT) method is introduced to obtain the tight outer approximation of the feasible region. Though this method can obtain less conservative interval solutions than previous methods, the results in any case are still supersets of the exact intervals.

The method used in [20] belongs to a class of methods named convex relaxation. The general idea is to enlarge the feasible sets by lifting the non-convex problems into a higher dimensional space where they are convex [157]. Then the convex problems are solved and the solutions are projected back onto the original space. These methods generally provide lower (upper) bounds for min. (max.) problems. When the projected solutions happen to lie in the feasible sets of the original problems, they obtain the global solutions of the original problems. This idea attracts much attention in optimal power flow (OPF) problems. After the seminal paper [154] making the important observation that SDP relaxation can find the global optimal solutions for several benchmark systems, much work has been devoted to the exactness conditions [158, 159] and implementation issues [160, 161] of convex, especially SDP relaxation of OPF. As conventional SDP relaxation is not always exact for the OPF problems, people are still seeking for tighter convex relaxations. Recent development in polynomial optimization provides a moment-SOS (sum of squares) approach to construct a hierarchy of SDP relaxations whose optima asymptotically [162] and often finitely [163] converge to the global optimum of the original non-convex problem. This approach has already been applied to OPF problems and obtained the global solutions of some cases where the conventional SDP method failed [164–166]. Due to the similarity in problem formulation, moment-SOS approach has the potential to tackle IPF problem as well.

The contribution of this chapter is applying sparsity-exploiting moment-SOS approach to IPF problems. Previous methods for IPF, in any case, can only obtain either outer approximation (e.g. IA, AA and LP relaxation methods) or inner approximation (e.g. Monte Carlo method) of the true intervals. In contrast, the moment-SOS approach can obtain the theoretically guaranteed exact interval solutions on small-scale systems. For larger systems, moment-SOS approach can also be employed to further tighten the interval solutions obtained by other methods, e.g. LP relaxation method [20], when high accuracy bounds are needed.

The rest of this chapter is organized as follows. Mathematical notations and definitions used in the manuscript are described in section 6.2. The polynomial optimization formulation of IPF problems is presented in section 6.3. Section 6.4 introduces the moment relaxation of IPF problems and section 6.4 describes the sparse techniques in moment relaxation. Case studies are reported in section 6.5. Finally, section 6.6 draws conclusions and gives suggestions on future research.

Notation and Definition

Let $\mathbb{R}$ be the set of real numbers, and $\mathbb{N}$ the set of nonnegative integers. $\mathbb{R}^n$ and $\mathbb{N}^n$ denote the sets of n-dimensional real and integer vectors, respectively. Boldface lower case letter $\boldsymbol{a}$ denotes a real vector with lowercase letter a_i denoting its i^{th} scalar element. $\mathbb{N}^n_d = \{\boldsymbol{\alpha} \in \mathbb{N}^n | \sum_{i=1}^n \alpha_i \leq d\}$ for $d \in \mathbb{N}$. Parentheses are used to construct vectors from comma separated

lists as $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k) = (\boldsymbol{x}_j)_{1\leq j\leq k} = [\boldsymbol{x}_1^T, \ldots, \boldsymbol{x}_k^T]^T$. Calligraphic uppercase letter $\mathcal{A}$ denotes a index set and $|\mathcal{A}|$ is its cardinality, i.e. number of elements. Matrix $A \succcurlyeq \mathbf{0}$ means that A is positive semidefinite. $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

$\mathbb{R}[\boldsymbol{x}]$ denotes the set of real valued polynomials in $x_i, i = 1, \ldots, n$. Each polynomial $f \in \mathbb{R}[\boldsymbol{x}]$ is represented as $f(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha}\in\mathcal{F}} c(\boldsymbol{\alpha})\boldsymbol{x}^{\boldsymbol{\alpha}}$ for a finite set $\mathcal{F} \in \mathbb{N}^n$ and some real numbers $c(\boldsymbol{\alpha})$ where $\boldsymbol{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_n^{\alpha_n}$ and $\mathcal{F}$ denotes the set of index vector related to its monomials. The set $\mathcal{F}$ is also called the set of supports denoted as $\text{supp}(f)$. The degree of $f \in \mathbb{R}[\boldsymbol{x}]$ is denoted by $\deg(f) = \max\{\sum_{i=1}^n \alpha_i | \boldsymbol{\alpha} \in \text{supp}(f)\}$.

A clique of a undirected graph $G(\mathcal{N}, \mathcal{E})$ is a subset of vertices such that its induced subgraph is complete, i.e. there is a edge between every two distinct vertices. A maximal clique is a clique that is not a proper subset of another clique. A chord is any edge joining two nonconsecutive vertices of a cycle. An undirected graph is chordal if every cycle of length greater than three has a chord. A graph $G(\mathcal{N}, \overline{\mathcal{E}})$ is a chordal extension of $G(\mathcal{N}, \mathcal{E})$ if it is a chordal graph and $\mathcal{E} \subseteq \overline{\mathcal{E}}$.

6.2 IPF Problem Formulation

In this section, IPF problems are formulated as polynomial optimization problems. Consider a power network with n_b buses and n_l lines. $\mathcal{N}$ denotes the set of bus indexes and $\mathcal{L}$ is the set of line indexes. $\mathcal{N}_{pv}$ and $\mathcal{N}_{pq}$ denote the index sets for PV and PQ buses, respectively. To obtain a polynomial formulation, bus voltages are expressed in rectangular coordinates. Let e_k and f_k denote the real and imaginary parts of complex bus voltage of bus k. Further define $\boldsymbol{v}_k = (e_k, f_k)$ and $\boldsymbol{v} = (\boldsymbol{v}_k)_{k\in\mathcal{N}}$. Without loss of generality, the first bus is assumed to be the reference bus and set $f_1 = 0$. The $(j,k)^{th}$ element of the network admittance matrix is denoted by $g_{jk} + \mathrm{j}b_{jk}$. Due to the uncertainty of loads and generations, the exact values of power injections are unknown whereas the estimated intervals are available. When the power injections stay in the estimated intervals, the power flow solutions must also be within certain intervals. IPF hence aims to obtain the the upper and lower bounds for power flow solutions, formally stated as the following optimization problem:

$$\min_{\boldsymbol{v}}(\max)\ f(\boldsymbol{v}) \qquad \text{subject to} \tag{6-1a}$$

$$\underline{P}_k \leq e_k \sum_{j=1}^{n_b}(g_{jk}f_j - b_{jk}e_j) + f_k \sum_{j=1}^{n_b}(g_{jk}e_j + b_{jk}f_j) \leq \overline{P}_k, \quad \forall k \in \mathcal{N}_{pv} \cup \mathcal{N}_{pq} \tag{6-1b}$$

$$\underline{Q}_k \leq f_k \sum_{j=1}^{n_b}(g_{jk}f_j - b_{jk}e_j) - e_k \sum_{j=1}^{n_b}(g_{jk}e_j + b_{jk}f_j) \leq \overline{Q}_k, \quad \forall k \in \mathcal{N}_{pq} \tag{6-1c}$$

$$e_k^2 + f_k^2 = U_k^2, \qquad \forall k \in \mathcal{N}_{pv} \tag{6-1d}$$

$$e_1 = U_1,\ f_1 = 0 \tag{6-1e}$$

$$e_k^2 + f_k^2 \geq V_m^2, \qquad \forall k \in \mathcal{N} \tag{6-1f}$$

where the objective $f(\boldsymbol{v})$ can be voltage magnitude (V.M.) $e_k^2 + f_k^2$, voltage angle (V.A.) f_k/e_k, line active power (A.P.)

$$g_{ij}(e_i^2 + f_i^2) - (e_i g_{ij} e_j - e_i b_{ij} f_j + f_i g_{ij} f_j + f_i b_{ij} e_j) \tag{6-2}$$

and line reactive power (R.P.)

$$- b_{ij}(e_i^2 + f_i^2) - (f_i g_{ij} e_j - f_i b_{ij} f_j - e_i g_{ij} f_j - e_i b_{ij} e_j). \tag{6-3}$$

Note that the rational function f_k/e_k is used for voltage angle because there is a bijection between f_k/e_k and $\operatorname{actan}(f_k/e_k)$ and the moment-SOS approach discussed later allows for an elegant way to deal with rational objectives. It is well known that there exist some low voltage solutions to the power flows equations which are strongly related to voltage instability [167]. The inequality constraint (6-1f) is thereby added to exclude such unrealistic operational points. Problem (6-1) is a non-convex optimization problem for which conventional interior point method can only guarantee a local minimum. Similar to what happened to OPF problems, global solutions can be obtained by proper convex relaxations.

Problem (6-1) is a direct extension of conventional power flow (PF) problem. Loads and generations are modelled as nodal power injections. Buses are classified into three types, e.g. PQ, PV and slack buses [168]. Active and Reactive power injection limits are set for PQ buses. PV buses entail active power injection limits and fixed voltage magnitudes. Slack bus is taken as voltage reference point with fixed complex voltage. Similar to PF, after solving IPF (6-1), if the reactive power injections at some PV buses exceed their available reactive power upper (lower) limits, those PV buses will be converted to PQ buses by just setting the reactive power injections as the upper (lower) limits. Then the problem is solved again. This process is repeated until no reactive power injection limits at PV buses are violated. Flowchart of this procedure is given in [20].

6.3 Moment Relaxations of IPF Problems

Moment-SOS approach can be understood from two viewpoints. The first viewpoint is based on the sum-of-squares (SOS) representation of nonnegative polynomials [169]. The second one considers polynomial optimization problems as generalized moment problems [162] [170]. These two viewpoints are actually a primal and dual pair in the sense of a generalize lagrangian function [171] and their generated SDPs also maintain a primal-dual relationship. In practice, the moment viewpoint is often adopted because it offers easily checkable conditions to certify the exactness of the SDP relaxations.

Consider a compact form of problem (6-1): minimizing a polynomial objective func-

tion $f(\boldsymbol{v})$ (6-1a) over a compact set $\mathbb{K} \subset \mathbb{R}^n$ defined by a tuple of polynomial equalities $\{h_m(\boldsymbol{v}) = 0\}_{m\in\mathcal{E}}$ (6-1d,6-1e) and inequalities $\{g_n(\boldsymbol{v}) \geq 0\}_{n\in\mathcal{I}}$ (6-1b,6-1c,6-1f). We first explain the method for polynomial objective functions and the modification needed for rational objective functions will be mentioned later. The moment approach is based on the observation that problem

$$\min_{\boldsymbol{v}\in\mathbb{K}} f(\boldsymbol{v}) \tag{6-4}$$

is equivalent to

$$\min_{\mu\in\mathscr{M}(\mathbb{K})_+} \int_{\mathbb{K}} f d\mu \text{ s.t.} \int_{\mathbb{K}} d\mu = 1 \tag{6-5}$$

where $\mathscr{M}(\mathbb{K})_+$ is the set of all non-negative measures on $\mathbb{K}$. Let f^* and ρ^* be the minimums of the problem (6-4) and (6-5), respectively. The equivalence of the above two problems is trivial: $f(x) \geq f^*$ on $\mathbb{K}$ implies $\int_{\mathbb{K}} f d\mu \geq f^*$, therefore $f^* \leq \rho^*$; take $\mu = \delta_{x^*}$ which is the Dirac measure at the minimizer of the first problem, then $\rho^* \leq \int_{\mathbb{K}} f d\delta_{x^*} = f^*$. Note that problem (6-4) is generically nonlinear and non-convex while problem (6-5) is always convex though infinite dimensional. Therefore problem (6-5) is the main focus in the sequel.

The restrictive structure of polynomials enable us characterize the measure on $\mathbb{K}$ with a infinite sequence of moments, i.e. $\boldsymbol{y} = (y(\boldsymbol{\alpha}))_{\boldsymbol{\alpha}\in\mathbb{N}^{2n_b}}$ with $y(\boldsymbol{\alpha}) = \int_{\mathbb{K}} \boldsymbol{v}^{\boldsymbol{\alpha}} d\mu$ for some $\mu \in \mathscr{M}(\mathbb{K})_+$. Therefore, the unknown measure in problem (6-5) can be replaced by its sequence of moments $\boldsymbol{y}$. Define the Riesz linear functional $L_{\boldsymbol{y}} : \mathbb{R}[\boldsymbol{v}] \mapsto \mathbb{R}$ associated with a moment sequence $\boldsymbol{y}$ as $f(\boldsymbol{v}) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^n} c(\boldsymbol{\alpha})\boldsymbol{v}^{\boldsymbol{\alpha}} \mapsto L_{\boldsymbol{y}}(f) = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^n} c(\boldsymbol{\alpha}) y(\boldsymbol{\alpha})$. Thus, problem (6-5) is transformed to

$$\min_{\boldsymbol{y}\in\mathbb{R}^\infty} L_{\boldsymbol{y}}(f) \text{ s.t.} \begin{array}{l} \exists \mu \in \mathscr{M}(\mathbb{K})_+, y(\boldsymbol{\alpha}) = \int_{\mathbb{K}} \boldsymbol{v}^{\boldsymbol{\alpha}} d\mu, \forall \boldsymbol{\alpha} \in \mathbb{N}^n \\ y(\boldsymbol{0}) = 1 \end{array} \tag{6-6}$$

The condition under which a given sequence $\boldsymbol{y}$ is the moment sequence of some positive measure on $\mathbb{K}$ has long been studied known as the $\mathbb{K}$-moment problem [170]. To state this condition, the definitions of moment matrix and localizing matrix are needed. The moment matrix $M_d(\boldsymbol{y})$ associated with $\boldsymbol{y}$ is the real symmetric matrix with rows and columns indexed in certain monomial basis $(\boldsymbol{v}^{\boldsymbol{\alpha}})$ with entries $M_d(\boldsymbol{y})(\boldsymbol{\alpha}, \boldsymbol{\beta}) = L_{\boldsymbol{y}}(\boldsymbol{v}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}) = y(\boldsymbol{\alpha}+\boldsymbol{\beta})$, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^n_d$. Similarly, the localizing matrix $M_d(g\boldsymbol{y})$ associated with $\boldsymbol{y}$ and polynomial $g \in \mathbb{R}[\boldsymbol{v}]$ is the real symmetric matrix with rows and columns indexed in monomial basis $(\boldsymbol{v}^{\boldsymbol{\alpha}})$ with entries $M_d(g\boldsymbol{y})(\boldsymbol{\alpha}, \boldsymbol{\beta}) = L_{\boldsymbol{y}}(g(\boldsymbol{v})\boldsymbol{v}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}) = \sum_{\boldsymbol{\gamma}} g_{\boldsymbol{\gamma}} y(\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma})$, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}^n_d$. It turns out that, under mild assumption, sequence $\boldsymbol{y}$ is a moment sequence for some positive measure on $\mathbb{K}$ if and only if $\forall d \in \mathbb{N}$, $M_d(\boldsymbol{y}) \succcurlyeq \boldsymbol{0}$, $M_d(h_m\boldsymbol{y}) = \boldsymbol{0}$, $\forall m \in \mathcal{E}$, $M_d(g_n\boldsymbol{y}) \succcurlyeq \boldsymbol{0}$, $\forall n \in \mathcal{I}$ [170]. To make it numerically trackable, it is necessary to relax it to a finite dimensional problem by limiting the order of monomials involved. Let $d_f = \lceil deg(f)/2 \rceil$, $d_{h_m} = \lceil deg(h_m)/2 \rceil$, $\forall m \in \mathcal{E}$ and $d_{g_n} = \lceil deg(g_n)/2 \rceil$, $\forall n \in \mathcal{I}$. For a fixed $d \geq \max\{d_f, \{d_{h_m}\}_{m\in\mathcal{E}}, \{d_{g_n}\}_{n\in\mathcal{I}}\}$, it results in the

following SDP:

$$\min_{\boldsymbol{y}\in\mathbb{R}^{\binom{2n_b+2d}{2d}}} L_{\boldsymbol{y}}(f) \qquad \text{subject to} \tag{6-7a}$$

$$M_d(\boldsymbol{y}) \succcurlyeq \mathbf{0} \tag{6-7b}$$

$$M_{d-d_{h_m}}(h_m\boldsymbol{y}) = \mathbf{0} \qquad \forall m \in \mathcal{E} \tag{6-7c}$$

$$M_{d-d_{g_n}}(g_n\boldsymbol{y}) \succcurlyeq \mathbf{0} \qquad \forall n \in \mathcal{I} \tag{6-7d}$$

$$y(\mathbf{0}) = 1 \tag{6-7e}$$

Problem (6-7) serves as a finite dimensional relaxation and its optimum is a lower bound of that of problem (6-5). By increasing the relaxation order d, it leads to a hierarchy of semidefinite programs whose optima asymptotically converge to that of problem (6-5) [162]. Moreover, finite convergence happens in generic problems [163].

In this chapter, the original polynomial optimization problem is assumed to have unique global optimum. This assumption is reasonable in practical IPF problems because the feasible sets of IPF problems are compact, thus any small random perturbation of the objective function makes its solution unique. The condition under which the order-d moment relaxation (6-7) is exact can then be stated as rank $M_d(\boldsymbol{y}^*) = 1$ where $\boldsymbol{y}^*$ is the optimal solution of order-d SDP realxation (6-7). If this condition is satisfied, the spectral decomposition of the diagonal block of $M_d(\boldsymbol{y}^*)$ related to the second-order terms, i.e. $L_{\boldsymbol{y}^*}(\boldsymbol{v}\boldsymbol{v}^T)$ yields the global optimal solution of (6-4) i.e. $\boldsymbol{v}^* = \sqrt{\lambda_1}\boldsymbol{\eta}_1$ where λ_1 is the non-zero eigenvalue and $\boldsymbol{\eta}_1$ is the corresponding eigenvector. Even if the rank-1 condition is not strictly satisfied, the above formula produces an approximate solution with λ_1 denoting the largest eigenvalue.

The whole method only needs minor modification to deal with rational objective functions. For $f(\boldsymbol{v}) = r(\boldsymbol{v})/s(\boldsymbol{v})$, modified problem (6-5) writes $\min_{\mu\in\mathscr{M}(\mathbb{K})_+} \int_{\mathbb{K}} r d\mu$ s.t. $\int_{\mathbb{K}} s d\mu = 1$. Accordingly, (6-7a) is modified to $L_{\boldsymbol{y}}(r)$ and (6-7e) is replaced by $L_{\boldsymbol{y}}(s) = 1$. In addition, the solution is extracted by $\boldsymbol{v}^* = \sqrt{\lambda_1}\boldsymbol{\eta}_1/\sqrt{\boldsymbol{y}^*(\mathbf{0})}$. This modification is based on method presented in [170].

6.4 Exploiting Sparsity in Moment Relaxations

IPF problem formulated as (6-1) is far from a generic polynomial problem but rather present some sparsity. If sparsity is properly exploited, it will lead to more efficient algorithm. In this section, we introduce the sparse moment relaxation developed in [172, 173] and applied to OPF in [164, 166]. Roughly speaking, this sparse moment relaxation rests on the observation that each equality or inequality constraint only involves a small subset of variables and the objective function can also be partitioned into polynomials involving only these small subsets of variables. Then under proper restriction on these subsets, the matrix equalities and inequalities

in moment relaxation (6-7) can be decomposed into several smaller parts.

The sparsity in IPF problem (6-1) is precisely defined as follows. The sparsity of (6-1) is described in terms of an $n_b \times n_b$ symmetric symbolic matrix $\boldsymbol{R}$. Its element R_{ij} is 1 if and only if either (i) $i = j$, or (ii) $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ appear simultaneously in p_k or q_k for some $k \in \mathcal{N}$. All other elements of $\boldsymbol{R}$ are zeros. Note that condition (ii) above actually means either bus i and bus j are connected or they both connect to a third bus. IPF problem (6-1) is sparse if matrix $\boldsymbol{R}$ is sparse. To reduce the size of moment relaxation (6-7), the set of indexes of decision variables $\mathcal{N}$ needs to be partitioned into several possibly overlapping subsets according to the sparsity pattern of $\boldsymbol{R}$. The sparsity pattern graph of $\boldsymbol{R}$ is a undirected graph $G(\mathcal{N}, \mathcal{E})$ with $\mathcal{E} = \{(i,k) | i,k \in \mathcal{N}, i < k, R_{ik} = 1\}$.

Let $\mathcal{N}$ be the union $\cup_{i=1}^{p} \mathcal{N}_i$ of p possibly overlapping subsets $\mathcal{N}_i, i = 1, \ldots, p$. Define the sets of supports $\mathcal{A}^{\mathcal{N}_i} = \{\boldsymbol{\alpha} \in \mathbb{N}^{2n_b} | \alpha_{2k-1} = 0, \alpha_{2k} = 0, \forall k \notin \mathcal{N}_i\}$ and $\mathcal{A}_{2d}^{\mathcal{N}_i} = \{\boldsymbol{\alpha} \in \mathbb{N}_{2d}^{2n_b} | \alpha_{2k-1} = 0, \alpha_{2k} = 0, \forall k \notin \mathcal{N}_i\}$. Assume the partition $\mathcal{N} = \cup_{i=1}^{p} \mathcal{N}_i$ satisfies

$$\forall k \in \mathcal{N},\ \exists\, \mathcal{N}_i,\ \text{supp}(p_k) \subseteq \mathcal{A}^{\mathcal{N}_i} \text{ and } \text{supp}(q_k) \subseteq \mathcal{A}^{\mathcal{N}_i}. \tag{6-8}$$

Subsequently, $\mathcal{N}$ can be partioned into p disjoint sets $\mathcal{J}_i \subseteq \mathcal{N}_i, i = 1, \ldots, p$ such that $\forall 1 \leq i \leq p, \forall k \in \mathcal{J}_i, \text{supp}(p_k) \subseteq \mathcal{A}^{\mathcal{N}_i}$ and $\text{supp}(q_k) \subseteq \mathcal{A}^{\mathcal{N}_i}$, which can be done by assigning k to $\mathcal{J}_h$ such that $\mathcal{N}_h$ is the smallest set among all $\mathcal{N}_i$ satisfying $\text{supp}(p_k) \subseteq \mathcal{A}^{\mathcal{N}_i}$ and $\text{supp}(q_k) \subseteq \mathcal{A}^{\mathcal{N}_i}$. Observe that the set $\mathcal{C}_k = \{i \in \mathcal{N} |\ p_k \text{ involves } \boldsymbol{v}_i\}$ forms a clique of $G(\mathcal{N}, \mathcal{E})$ by definition of $\boldsymbol{R}$. Hence $\mathcal{C}_k$ is contained in one maximal clique of $G(\mathcal{N}, \mathcal{E})$. In light of this, assumption (6-8) can be satisfied if each $\mathcal{N}_i$ is a maximal clique of $G(\mathcal{N}, \mathcal{E})$. In addition, assume that for every $i = 1, \ldots, p-1$,

$$\exists\, r \in \{1, \ldots, i\},\ \mathcal{N}_{i+1} \cap (\mathcal{N}_1 \cup \cdots \cup \mathcal{N}_i) \subseteq \mathcal{N}_r. \tag{6-9}$$

Assumption (6-9) is known as the running intersection property in graph theory, which is satisfied when $\mathcal{N}_1, \ldots, \mathcal{N}_p$ are the maximal cliques of a chordal graph [174]. The following strategy is proposed in [172] to obtain a partition of $\mathcal{N}$ satisfying both (6-8) and (6-9). First, generate a chordal extension $G(\mathcal{N}, \mathcal{E}')$ of the graph $G(\mathcal{N}, \mathcal{E})$. Then find all maximal cliques $\mathcal{N}_i, i = 1, \ldots, p$ of $G(\mathcal{N}, \mathcal{E}')$. For implementation convenience, the above strategy can be realized through Cholesky factorization of $\boldsymbol{R} + \delta \boldsymbol{I}$ after a symmetric approximate minimum degree ordering, and the sparse pattern of the Cholesky factor defines the variable partition.

Under condition (6-8) and (6-9), the following sparse moment relaxation is well-defined with its optimum also converging to the global optimum of the original problem [170]:

$$\min_{\boldsymbol{y} \in \mathbb{R}^{\binom{2n_b+2d}{2d}}} L_{\boldsymbol{y}}(f) \qquad \text{subject to} \tag{6-10a}$$

$$M_{d_i}(\boldsymbol{y}, \mathcal{N}_i) \succcurlyeq \boldsymbol{0} \quad \forall 1 \leq i \leq p \tag{6-10b}$$

$$M_{d_i - d_{h_m}}(h_m \boldsymbol{y}, \mathcal{N}_i) = \boldsymbol{0} \quad \forall 1 \leq i \leq p, \forall m \in \mathcal{E}_i \tag{6-10c}$$

$$M_{d_i-d_{g_n}}(g_n\boldsymbol{y},\mathcal{N}_i)\succcurlyeq\mathbf{0}\quad\forall 1\leq i\leq p,\forall n\in\mathcal{I}_i \tag{6-10d}$$

$$y(\mathbf{0})=1 \tag{6-10e}$$

where $M_d(\boldsymbol{y},\mathcal{N}_i)$ is the moment submatrix obtained from $M_d(\boldsymbol{y})$ by retaining those rows and columns indexed by $\boldsymbol{\alpha}\in\mathcal{A}_d^{\mathcal{N}_i}$; similarly, $M_d(h_m\boldsymbol{y},\mathcal{N}_i)$ and $M_d(g_n\boldsymbol{y},\mathcal{N}_i)$ are localizing submatrices obtained from $M_d(h_m\boldsymbol{y})$ and $M_d(g_n\boldsymbol{y})$ by retaining those rows and columns indexed by $\boldsymbol{\alpha}\in\mathcal{A}_d^{\mathcal{N}_i}$; $\mathcal{E}_i$ (resp. $\mathcal{I}_i$) denotes the index set of inequality (resp. equality) constraints (6-1b,6-1c,6-1f) (resp. (6-1d,6-1e)) with $k\in\mathcal{J}_i$. The sparse moment relaxation is exact if rank $M_{d_i}(\boldsymbol{y},\mathcal{N}_i)=1,\forall 1\leq i\leq p$. The voltage $\boldsymbol{v}_k$ can then be extracted from the spectral decomposition of the diagonal block corresponding to second-order monomials in $M_{d_i}(\boldsymbol{y},\mathcal{N}_i)$ where $k\in\mathcal{J}_i\subseteq\mathcal{N}_i$. The sizes of positive semidefinite constraints in (6-10) are considerably smaller than those of (6-7) due to the sparsity of $\boldsymbol{R}$. Hence it can be solved more efficiently and applied to cases larger than (6-7).

6.5 Case Studies

To demonstrate the applicability and analyze the performance of moment-SOS approach to IPF problems, numerical studies are conducted on several IEEE standard systems. All the test data is extracted from MATPOWER 4.1 [116]. Moment-SOS approach is programmed in MATLAB with YALMIP [175] as the modeling tool and Mosek [176] as the solver. The program runs on a Win8 PC with a 3.0 GHz CPU with 8GB RAM. For brevity, PV-PQ bus type switching is not conducted in the followed case studies.

6.5.1 Exact Global Solutions on Small Cases

The sparse moment relaxation discussed above is directly applied to small-size IEEE 6-bus, 9-bus, and 14-bus systems with $\pm10\%$ uncertainty on the loads to check whether moment-SOS approach can obtain the exact global interval power flow solutions. Two metrics are used to measure the global optimality of the solutions. The first metric is the smallest ratio between the largest and the second largest eigenvalues of all the moment matrices $M_{d_i}(\boldsymbol{y},\mathcal{N}_i)$. The second metric is the largest violation of the extracted solution $\boldsymbol{v}^*$ to the constraints (6-1b)$\sim$(6-1f). The moment-SOS approach attains the global optimal solution if the first metric is large enough which certifies the satisfaction of the rank-1 condition and the second metric is small enough which certifies the feasibility of the extracted solutions.

Some numerical results are reported in Table 6-1 where column 3 and 4 show the extracted interval solutions, column 5 is the smallest value of the first metric related to the four problems in the same row, column 6 shows the largest value of the second metric related to the four problems in the same row, and the last column shows the relaxation order needed to achieve these results. The notations V.M., V.A., A.P. and R.P. denote the bus voltage magnitude, the

bus voltage angle, the line active power and the line reactive power, respectively. For all the problems shown in Table 6-1, the values of the first metric are larger than 10e7 and the values of the second metric are smaller than 10e-6, which numerically certifies the global optimality of the extracted solutions. All the problems can be solved globally with the second-order moment relaxations. A small portion of problems can be solved with the first-order moment relaxations. Interestingly, all the max. V.M. problems are solved with first-order relaxation, and some max. A.P. and R.P. problems are also solved with first-order relaxation.

It is worth emphasizing the functionality of constraint (6-1f). In the implementation, the value of V_m^2 is set to be 0.5 p.u.. For example, with constraint (6-1f), the lower bound of the V.M. of bus 5 on IEEE-9 system is 0.9679 p.u. obtained by second-order moment relaxation. If constraint (6-1f) is dropped, the lower bound of V.M. of this bus is 0.0787 which can be obtained by first-order moment relaxation with the first metric 7.3e5 and the second metric 2.8e-6. Due to the existence of multiple solutions of power flow equations, one of the low voltage solutions is obtained if constraint (6-1f) is not added to problem (6-1). In addition, the constraint (6-1f) is not explicitly active (all the voltage magnitudes are strictly larger than 0.5 p.u.) which shows the complex nonlinear nature of power flow equations.

The exact interval power flow solutions of IEEE-9 and IEEE-14 systems are also shown in Fig. 6-1 and Fig. 6-2. Some inconsistency is observed between our results and the results reported in paper [20] on IEEE-9 system. Since this chapter shares the same problem formulations with [20] (see eq. (6)$\sim$(10) in [20]), the interval solution obtained by LP relaxation in paper [20] should not be tighter than the theoretically guaranteed exact interval solution obtained by moment-SOS approach in this chapter. However, a brief comparison shows some results in [20] are even tighter than the results shown in Table 6-1 in this chapter. The possible reason is that the initial estimation of the solution interval in [20] is so aggressive that it excludes the real global solutions.

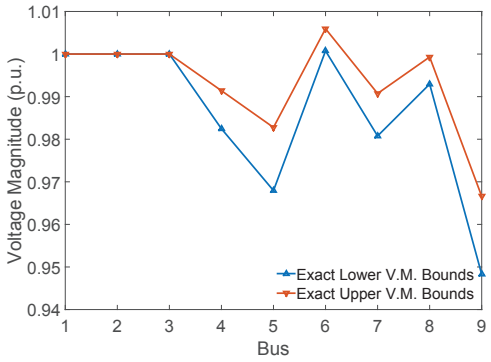
6.5.2 High Accuracy Solution on Larger Cases

Even though sparsity is exploited, second-order moment relaxation for systems with more than forty buses is still numerically intractable using current SDP solvers. As shown from Table I, the first moment relaxation is only exact on very limited portion of cases. Therefore, generally speaking, only approximate solutions can be obtained for larger systems at the current stage. However, OBBT discussed in [20] and selective application of high-order constraints proposed in [166] can help to obtain very tight interval solutions.

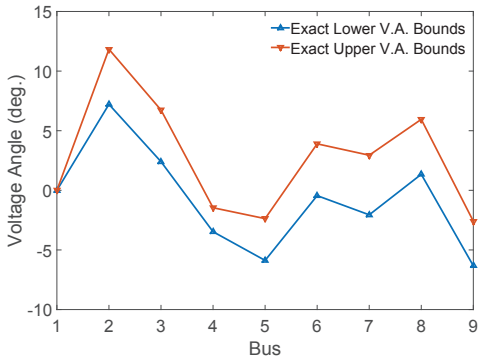
The moment relaxation discussed in this chapter can be used at the final stage of the algorithm framework discussed in [20], which means the LP relaxation based OBBT is firstly used to obtain the convex outer approximation of the feasible set (conceptually, SDP based OBBT can also be used, but it is too expensive). Then moment relaxation can be employed to

Table 6-1 Exact Interval Power Flow Solutions for 6, 9 and 14-Bus Systems under 10% Load Uncertainty

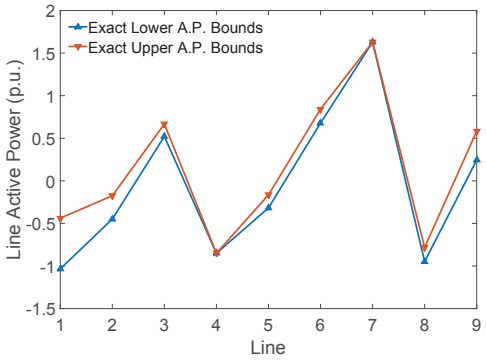
System	Bus/Line	V.M./A.P.	V.A./R.P.	min eig. ratio	max con. violation	relaxation order			
6-bus	4	[0.9819, 0.9967]	[-5.2053,-3.1978]	7.8e+8	7.4e-9	2nd	1st	2nd	2nd
	5	[0.9762, 0.9944]	[-6.5193,-4.0450]	1.6e+9	7.1e-9	2nd	1st	2nd	2nd
	6	[0.9973, 1.0114]	[-7.4406,-4.4832]	8.2e+8	9.4e-9	2nd	1st	2nd	2nd
	1-2	[-0.3712,-0.2055]	[0.0969, 0.1668]	4.2e+9	3.7e-9	2nd	1st	2nd	2nd
	2-6	[-0.3048,-0.2206]	[-0.1817,-0.1221]	4.6e+8	3.1e-8	2nd	2nd	2nd	1st
	4-5	[-0.0679,-0.0140]	[-0.0121, 0.0323]	2.5e+9	2.6e-9	2nd	2nd	2nd	2nd
9-bus	5	[0.9679, 0.9828]	[-5.8822,-2.1736]	1.5e+8	6.4e-8	2nd	1st	2nd	2nd
	7	[0.9801, 0.9908]	[-2.0765, 3.2822]	2.4e+8	3.4e-8	2nd	1st	2nd	2nd
	9	[0.9483, 0.9666]	[-6.3340,-2.3979]	1.5e+8	6.3e-8	2nd	1st	2nd	2nd
	1-4	[-1.0352,-0.4059]	[-0.3266,-0.1579]	8.5e+7	2.7e-7	2nd	1st	2nd	1st
	5-6	[0.5215, 0.6671]	[-0.0438, 0.0316]	1.7e+8	4.0e-8	2nd	2nd	2nd	2nd
	8-9	[-0.9477,-0.7825]	[-0.1670,-0.0877]	4.6e+8	3.5e-8	2nd	2nd	2nd	2nd
14-bus	4	[1.0144, 1.0208]	[-11.5329,-9.1053]	1.5e+8	1.7e-8	2nd	1st	2nd	2nd
	7	[1.0584, 1.0646]	[-14.9016,-11.8320]	1.1e+8	1.7e-8	2nd	1st	2nd	2nd
	13	[1.0478, 1.0529]	[-16.9197,-13.4119]	2.5e+7	4.9e-8	2nd	1st	2nd	2nd
	2-5	[-0.4582,-0.3725]	[-0.0387,-0.0230]	5.8e+7	4.5e-7	2nd	2nd	2nd	1st
	6-13	[-0.1989,-0.1561]	[-0.0842,-0.0603]	7.8e+7	3.6e-8	2nd	2nd	2nd	1st
	9-10	[-0.0745,-0.0300]	[-0.0553,-0.0291]	2.9e+8	9.5e-8	2nd	2nd	2nd	2nd



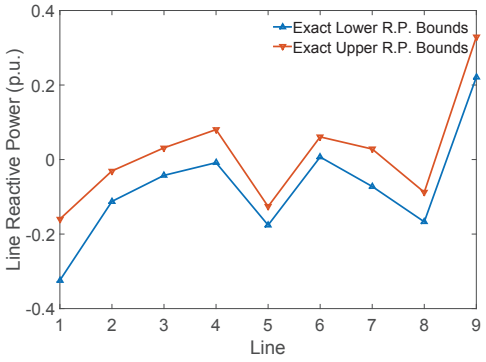
(a) Voltage magnitude



(b) Voltage angle



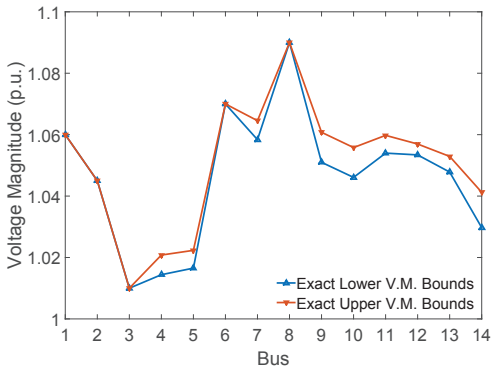
(c) Active power



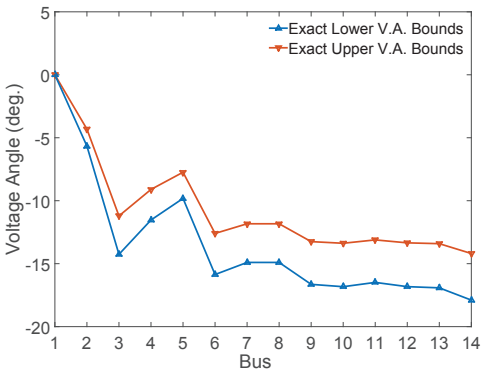
(d) Reactive power

Figure 6-1 Exact Interval Power Flow Bounds for IEEE-9 System

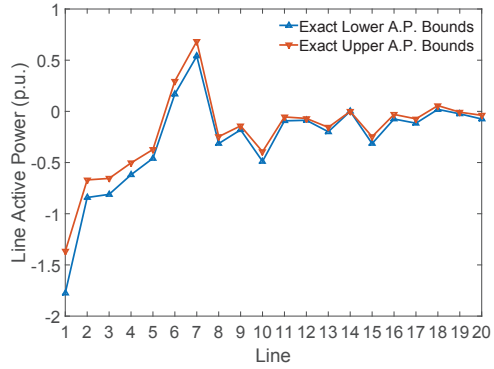
solve the final problem and obtain the lower and upper bounds of related objectives. To this end, the following constraints



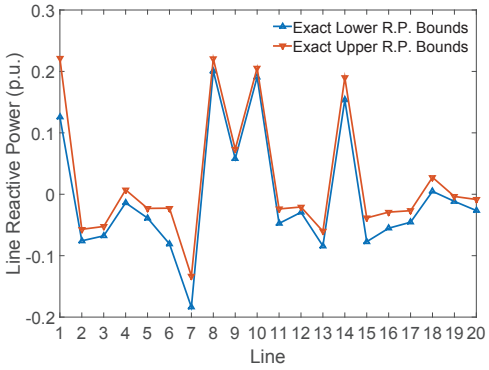
(a) Voltage magnitude



(b) Voltage angle



(c) Active power



(d) Reactive power

Figure 6-2 Exact Interval Power Flow Bounds for IEEE-14 System

$$\underline{v}_i \le v_i \le \overline{v}_i \tag{6-11a}$$
$$v_i v_j - \underline{v}_i v_j - \underline{v}_j v_i + \underline{v}_i \underline{v}_j \ge 0, \quad (i,j) \in \mathcal{L} \tag{6-11b}$$
$$v_i v_j - \overline{v}_i v_j - \overline{v}_j v_i + \overline{v}_i \overline{v}_j \ge 0, \quad (i,j) \in \mathcal{L} \tag{6-11c}$$
$$v_i v_j - \underline{v}_i v_j - \overline{v}_j v_i + \underline{v}_i \overline{v}_j \le 0, \quad (i,j) \in \mathcal{L} \tag{6-11d}$$
$$v_i v_j - \overline{v}_i v_j - \underline{v}_j v_i + \overline{v}_i \underline{v}_j \le 0, \quad (i,j) \in \mathcal{L} \tag{6-11e}$$

need to be added to the polynomial problem (6-1) before construct the moment relaxation.

The solutions obtained from the first-order moment relaxation satisfy or slightly violate most of the constraints (6-1b)∼(6-1f). Only very limited number of constraints are considerably violated. Therefore selectively apply high-order relaxation to these buses where large violation happens may significantly tighten the relaxation. To this end, each bus is associated with a relaxation order and the relaxation order of the maximal clique $\mathcal{N}_i$ is equal to the highest order of buses in $\mathcal{J}_i$. In this way, (6-10b) is constructed according to the order of the maximal clique and (6-10c)∼(6-10d) are constructed according to the order of the buses. In our implementation, for all bus $k \in \mathcal{J}_i \subseteq \mathcal{N}_i$ with $|\mathcal{N}_i| \le 8$, second-order relaxations are applied to two buses where the solution to first-order relaxation yields largest violation.

Table 6-2 Comparison of Bounds Solution on 57-Bus System

Problem	LP	ord-1 MR	ord-1.2 MR
4 V.M.	[0.9672,0.9941]	[0.9676,0.9816]	[0.9685,0.9816]
4 V.A.	[-11.1053,-3.3970]	[-11.0289,-4.4419]	[-11.0258,-4.4430]
7 V.M.	[0.9578,1.0119]	[0.9582,0.9863]	[0.9592,0.9863]
7 V.A.	[-13.3077,-1.4924]	[-13.1830,-3.0075]	[-13.1803,-3.0798]
3-4 A.P.	[-1.0530,-0.1512]	[-0.9706,-0.3440]	[-0.9503,-0.3461]
3-4 R.P.	[-0.5194,0.5746]	[-0.4804,0.1475]	[-0.4797,0.1472]
6-7 A.P.	[-0.1428,0.5148]	[0.0013,0.3664]	[0.0026,0.3336]
6-7 R.P.	[-0.4303,0.4179]	[-0.4036,0.0392]	[-0.3898,0.0389]

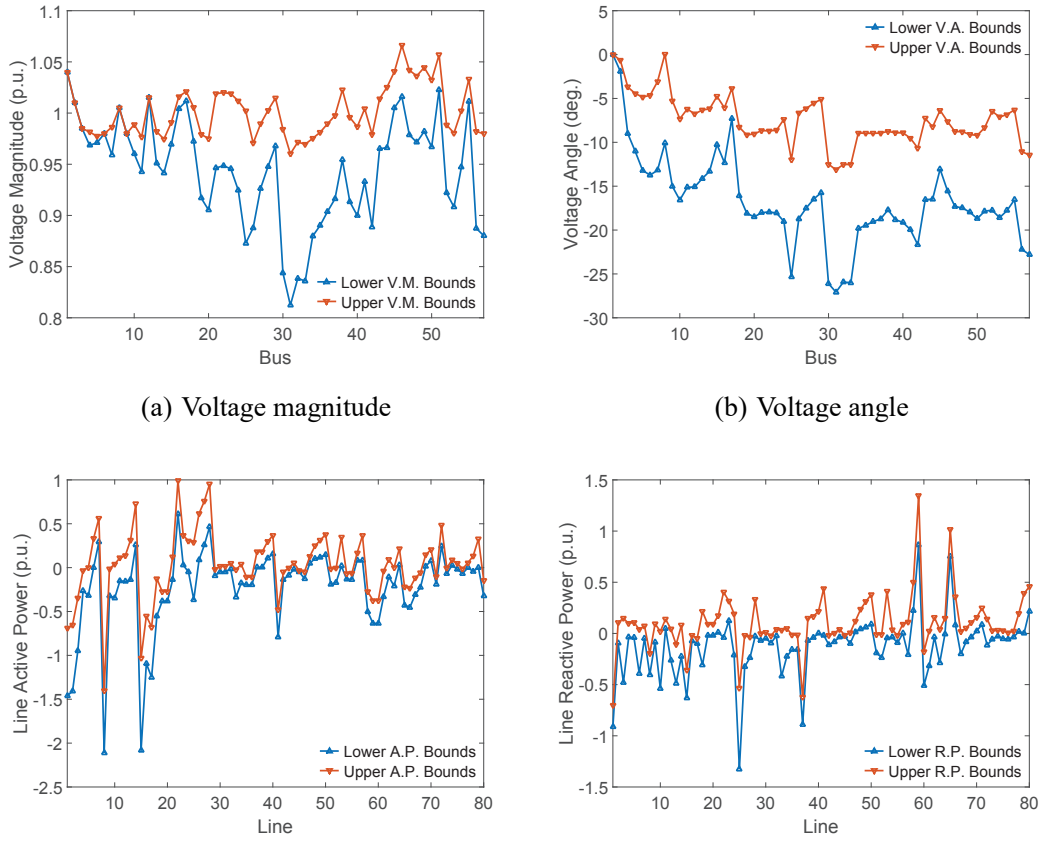
We have conducted comparative case studies among LP relaxation, first-order moment relaxation (denoted as ord-1 MR) and first-order moment relaxation with 2 second-order buses (denoted as ord-1.2 MR) on IEEE-57 and IEEE-118 systems with $\pm 10\%$ uncertainty on the loads. All the problems are solved by each method after exactly the same LP based OBBT procedure. Some numerical results are reported in Table 6-2 and Table 6-3. It clearly shows that ord-1 MR obtains tighter interval solutions than LP, and ord-1.2 MR obtains tighter interval solutions than ord-1 MR. The OBBT procedure really help to tighten the initial estimation of the solution interval so that the LP relaxation already obtains quite tight interval solutions on many cases. However, large improvement of ord-1 MR compared with LP is still observed on several cases, especially for the upper bounds of the line reactive power. For example, the upper bound for the reactive power of line 6-7 on IEEE-57 system obtained by LP is 0.4179 p.u., while that obtained by ord-1 MR is 0.0392. The upper bound of reactive power of line 4-5 on IEEE-118 system acquired by LP is 2.2667 compared to 0.2972 acquired by ord-1 MR. The improvement from ord-1 MR to ord-1.2 MR is not so significant compared with that from LP to ord-1 MR. Of course, increasing the number of second-order buses will further tighten the interval solutions but the marginal benefits will quickly diminish as shown in the comparison among LP, ord-1 MR and ord-1.2 MR. The interval power solutions obtained by ord-1.2 MR on IEEE-57 system is also shown in Fig. 6-3.

6.5.3 IEEE 300-bus System with Wind Power

The proposed method is also applied to IEEE 300-bus system with wind power uncertainties. Partial diagram of the system configuration is shown in Fig. 6-4. Three wind farms are installed at bus 225, 231 and 237 of the standard 300-bus system. Each wind farm has a power capacity of 80MW. The actual output power of each wind farm possesses $\pm 20\%$ uncertainty in terms of its capacity. The reactive power compensation devices and related control system in each wind farm always maintain a power factor of 1 at the connection point. Ord-1 MR is employed to obtain the interval solutions on several buses and lines which are heavily affected

Table 6-3 Comparison of Bounds Solution on 118-Bus System

Problem	LP	ord-1 MR	ord-1.2 MR
2 V.M.	[0.9612,0.9810]	[0.9612,0.9720]	[0.9675,0.9720]
2 V.A.	[-22.4996,-14.3718]	[-22.3451,-14.4066]	[-22.3443,-14.6189]
5 V.M.	[0.9927,1.0106]	[0.9930,1.0022]	[0.9938,1.0022]
5 V.A.	[-17.6181,-10.2282]	[-17.4044,-10.2582]	[-17.4041,-10.4862]
4-5 A.P.	[0.7346,1.3359]	[0.8414,1.2374]	[0.8883,1.1936]
4-5 R.P.	[-1.9190,2.2667]	[-1.8780,0.2912]	[-1.7711,0.2866]
5-6 A.P.	[-1.2543,-0.5347]	[-1.1742,-0.7266]	[-1.0978,-0.7631]
5-6 R.P.	[-0.2928,0.1884]	[-0.2020,0.1121]	[-0.2008,0.0995]



(a) Voltage magnitude (b) Voltage angle

(c) Active power (d) Reactive power

Figure 6-3 Interval Solutions by ord-1.2 MR on IEEE-57 System

by the uncertain output of wind farms. LP relaxation method is also introduced for comparison. Results are shown in Fig. 6-5 which demonstrates the proposed method can obtain much tighter interval solutions than LP relaxation method.

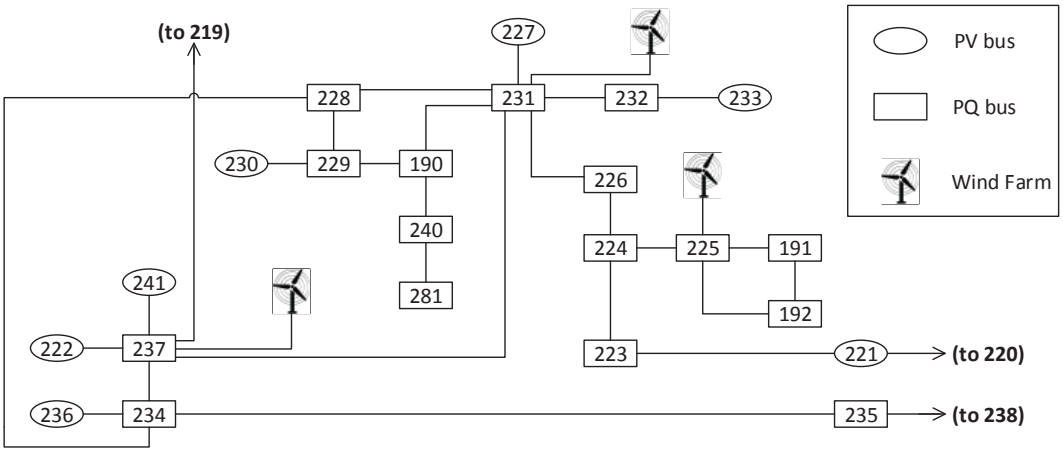
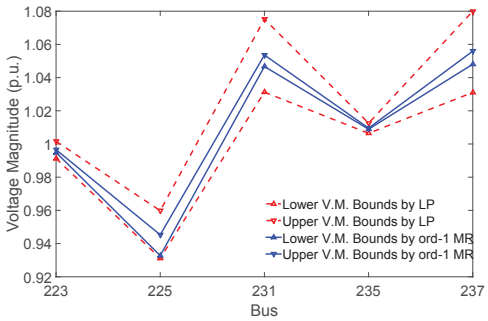
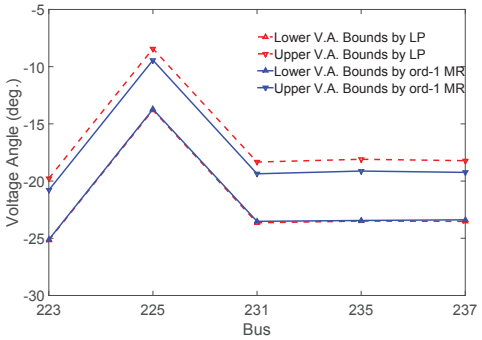


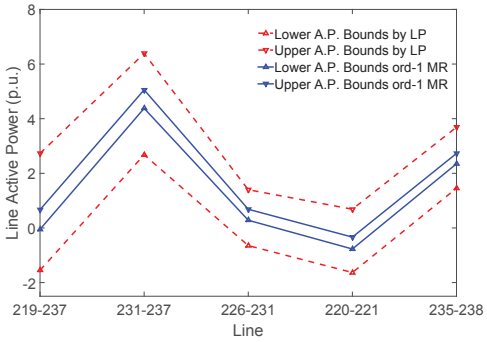
Figure 6-4 Partial Diagram of IEEE 300-bus System with Wind Farms



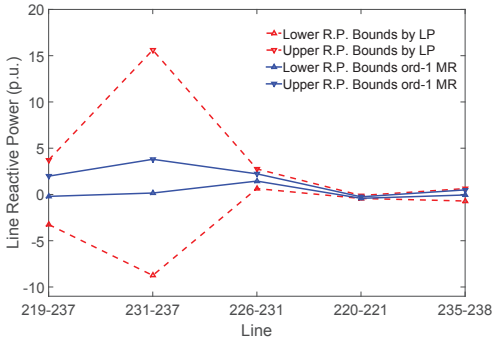
(a) Voltage magnitude



(b) Voltage angle



(c) Active power



(d) Reactive power

Figure 6-5 Comparison of ord-1 MR and LP on IEEE 300-bus System

6.5.4 Performance Analysis

Interval widths obtained by different relaxation methods are compared in Fig. 6-6, Fig. 6-7 and Fig. 6-8. Since the ord-2 MR obtains the exact interval solutions on IEEE-9 and IEEE-14 bus systems, the solid black line on Fig. 6-6 and Fig. 6-7 represent the exact interval widths. The ord-2 MR is numerically intractable for IEEE-57 system, so only LP, ord-1 MR and ord-1.2

Table 6-4 Solver Time Comparison (sec.)

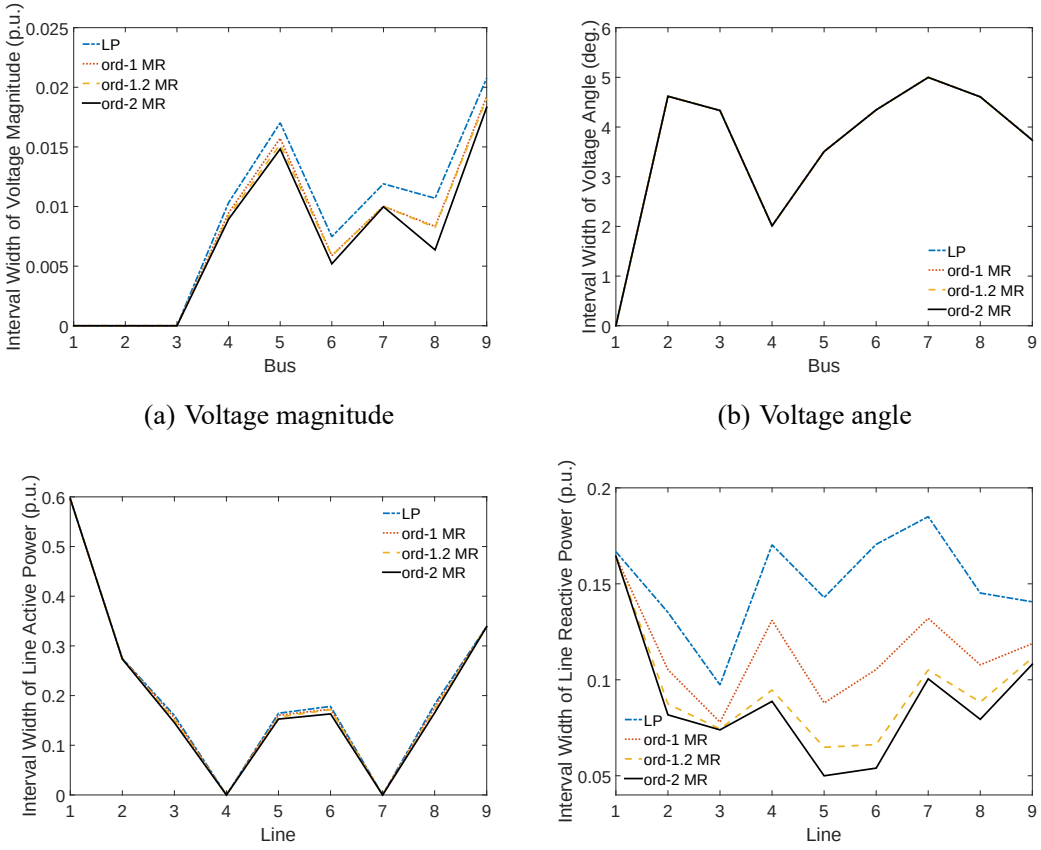
System	LP	ord-1 MR	ord-1.2 MR	ord-2MR
9-bus	0.17	0.2	2.3	3.64
14-bus	0.17	0.34	41.58	67.14
57-bus	0.27	2.47	30.98	/
108-bus	0.36	6.11	69.88	/

are drawn in Fig. 6-8. It is observed that the bound tightening effects of MRs compared with LP are generally very significant on line reactive power problems but much less significant on bus voltage angle problems. As shown in Fig. 6-6 and Fig. 6-7, the four lines of the voltage angles almost coincide with each other whereas the four lines of the line reactive power exhibit considerable differences. As system scale increases, the gaps between LP and MRs intensifies. As shown in Fig. 6-8, the LP relaxation yields very conservative interval solutions for some line active and reactive power problems. The difference between ord-1 MR and ord-1.2 MR is almost overwhelmed by the difference between LP and ord-1 MR, so the lines for ord-1 MR and the lines for ord-1.2 MR nearly coincide with each other in Fig. 6-8. The possible explanation for this phenomenon is that the severe non-linearity of line power functions may prevent LP relaxation makes a good approximation to optimization problem to be solved.

The bound tightening benefits of MRs compared with LP, of course, are not without costs. Typical solver time for a single problem needed for different methods is listed in Table 6-4. Since the sizes of the linear matrix inequality constraints increase dramatically as the system scale and relaxation order increase, the solver time needed for MRs increases significantly as the system scale and number of second-order buses increase.

Note that the SDP relaxation (equivalent to ord-1 MR [166]) for OPF problems is demonstrated applicable to systems with more than 3000 buses [161]. Therefore, the ord-1 MR should be considered as a practical remedy for LP relaxation if high accuracy interval solutions are needed. Moreover, for small-size systems, the ord-2 MR can give theoretically guaranteed exact interval solutions in stark contrast to the outer approximate given by IA, AA and LP methods and the inner approximation given by Monte Carlo simulation. When developing new methods, the results obtained by ord-2 MR are better standard results used for comparison than results given by Monte Carlo method often used in the literature.

Finally, We evaluate the effects of the sparsity-exploiting technique presented in Section V. Fig. 6-9 illustrates the partition of buses on IEEE 9-bus system. All buses are partitioned into 5 overlapping cliques and the largest clique contains 5 buses. Under this partition, the size of the largest SDP constraint for ord-2 MR (ord-1 MR) will decrease from 190×190 (19×19) to 66×66 (11×11). When this techniques is applied to IEEE 300-bus system, the size of the largest SDP constraint for ord-1 MR will decrease from 601×601 to 39×39. Such significant reduction



(a) Voltage magnitude (b) Voltage angle

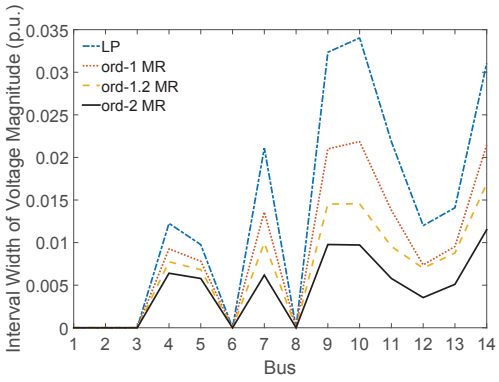
(c) Active power (d) Reactive power

Figure 6-6 Interval Width for IEEE-9 System Using Different Methods

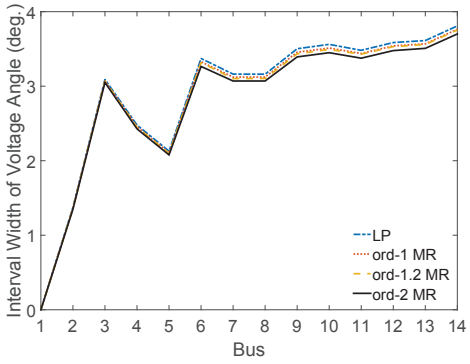
of constraint sizes will result in shorter solver time and higher tractability. The comparison of solver time needed before and after applying the sparisity exploiting technique is shown in Table 6-5 and Table 6-6. Since IEEE 6-bus system can only be partitioned into one clique, the sparsity-exploiting technique has no influence on its solution process and time. For other cases, it is shown that the sparsity-exploiting technique has brought orders-of-magnitude solver time saving for ord-1 MR on larger systems and for ord-2 MR even on small systems.

Table 6-5 Solver Time for ord-1 MR (sec.)

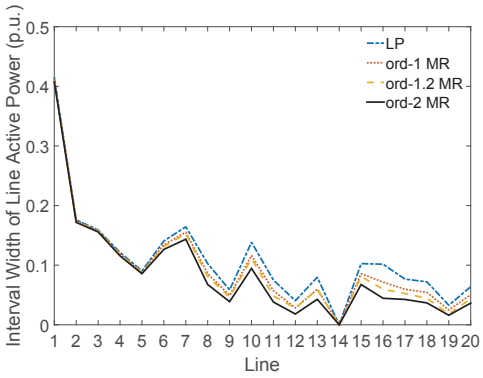
Exploit Sparsity	Systems					
	6	9	14	57	118	300
Yes	0.17	0.20	0.34	2.47	6.11	9.19
No	0.17	0.25	0.36	187	3346	intractable



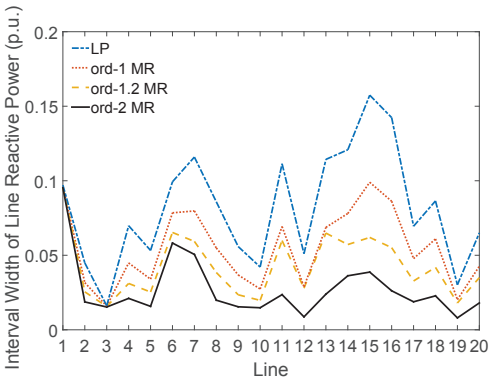
(a) Voltage magnitude



(b) Voltage angle



(c) Active power



(d) Reactive power

Figure 6-7 Interval Width for IEEE-14 System Using Different Methods

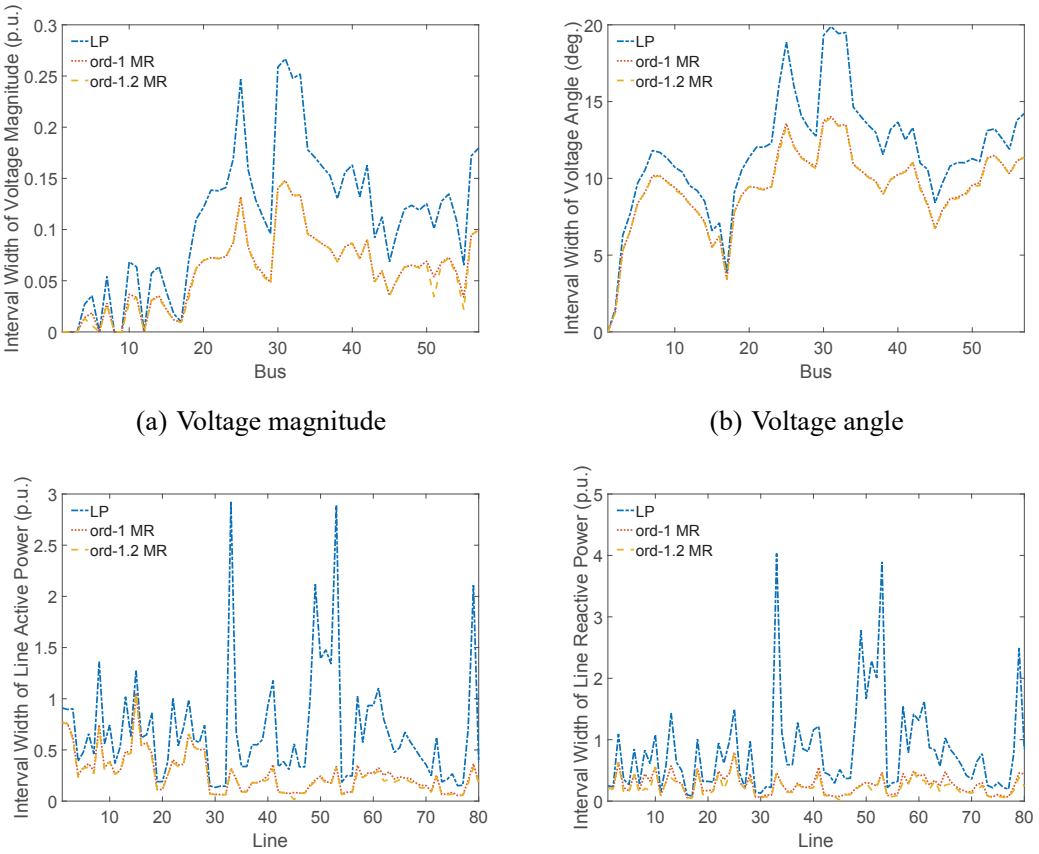
Table 6-6 Solver Time for ord-2 MR (sec.)

Exploit Sparsity	Systems		
	6	9	14
Yes	5.26	3.64	67.14
No	5.26	81.03	7385

6.6 Conclusions

In this chapter, moment-SOS approach is applied to interval power flow analysis which is formulated as polynomial optimization problems. Correlative sparsity of the problem formulation is exploited to improve numerical tractability and efficiency. Numerical studies on IEEE 6-bus, 9-bus and 14-bus systems demonstrate this approach is capable of obtaining exact interval solutions on small-scale systems. Moreover, this approach can significantly improve the interval solutions on larger systems based on numerical studies on IEEE 57-bus, 118-bus and 300-bus systems.

Note that the improvement of solution accuracy is at the cost of longer solver time. Therefore, future research will focus on two directions. First, seek for other approaches to construct



(a) Voltage magnitude (b) Voltage angle

(c) Active power (d) Reactive power

Figure 6-8 Interval Width for IEEE-57 System Using Different Methods

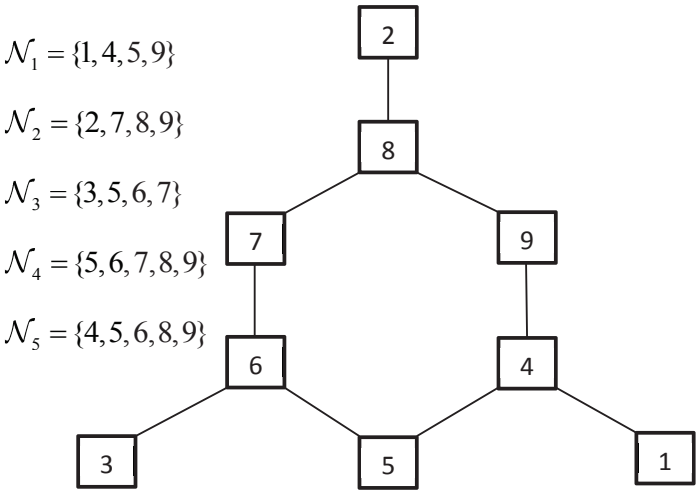


Figure 6-9 Partition of Buses on IEEE 9-bus System

convex relaxations for IPF problems where the complexity can be systematically controlled, so that it is flexible to choose suitable convex relaxation considering the tradeoff between accuracy and solver time. Second, speed up the computation by exploiting the separability in the

algorithm which allows for parallel computing.

7 Multi-Period OPF with Energy Storages and Renewable Sources: A Parallel Moment-SOS Approach

7.1 Introduction

Energy storage systems (ESS) are well-recognized as one of the most promising technologies to facilitate the integration of intermittent renewable energy since they provide the grid flexibility to manage non-dispatchable power sources. To maximize the benefit of ESSs, optimal power flow (OPF) technique should be considered in the operation of the power systems equipped with ESSs. Conventional single period OPF is nolonger suitable for a system with ESS because the operation of ESS is strongly coupled in time by charge/discharge dynamics. Therefore, multi-period OPF which takes into account charge/discharge dynamics as well as generator ramp-rate constraints is a more reasonable framework to guide the optimal operation of power systems with ESSs [75].

Multi-period OPF with ESSs has attracted much attention in recent years. Jabr *et al* incorporate robust optimization techniques into multi-peroid OPF framework to address the uncertainties of renewable generation [75, 105] where DC power flow equations are used to simplify the problem. DC flow based multi-peroid OPF is also employed in [91] to optimize storage allocation and portfolio. Since DC power flow neglects voltage magnitude and reactive power, its result might be unreasonable for practical operation, especially in distribution networks which have high R/X ratios and high variability of voltage magnitudes [92]. Thus, full AC formulations of multi-period OPF are used to address the optimal operation of distribution networks in [92] and [93]. But due to the non-convexity of the problem formulation and local search method used, only local optimality is guaranteed. By extending the seminar work of Lavaei and Low [65], semi-definite program (SDP) relaxations of AC flow based multi-period OPF are discussed in [94, 177, 178] which at least serves as a better approximation than DC flow based formulation, and in some circumstances achieves the global solution of the original non-convex problem.

Convex relaxation methods have attracted much attention in single-period optimal power flow (OPF) problems. After the seminal paper [65] making the important observation that SDP relaxations are exact for several benchmark systems, much work has been devoted to the exactness conditions [66, 67] and implementation issues [179, 180]. Since conventional SDP relaxation is not always exact for the OPF problems, tighter convex relaxations are needed to achieve global optimality. Moment-sos (sum of squares) approach is a systematic way to construct a hierarchy of SDP relaxations whose optima asymptotically converge to the global optimum of the original non-convex problem. This approach has already been applied to single-period

OPF and obtained the global solutions of some cases where the conventional SDP method failed [164–166].

In this chapter, the moment relaxation method is applied to the full AC multi-period OPF problems with energy storages and renewable sources. Since moment relaxations can be large-scale SDPs even for relatively small problems, a decomposition algorithm based on alternating direction method of multipliers (ADMM) is designed to improve the numerical tractability. The original time-correlated non-convex optimization problem is then decomposed into two subproblems. The first subproblem is non-convex but separable among different time slots, which can be solved in parallel among different time slots. In fact, for each time slot, it is a single-period OPF problem for which moment relaxation can be constructed and solved to obtain near optimal or even global optimal solutions. Moreover, sparsity is exploited to improve efficiency. The second subproblem is convex QP and separable among different buses, which can be solved by standard IPM solver in parallel.

7.2 Multi-Period OPF Problem Formulation

In this section, we formulate the multi-period OPF with energy storage and renewable generation as a nonlinear, nonconvex polynomial optimization problem.

Consider a power network represented by a connected undirected graph $G(\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ denotes the set of buses and $\mathcal{L} \subseteq \mathcal{N} \times \mathcal{N}$ denotes the set of lines. The set of conventional generator buses is denoted as $\mathcal{G} \subseteq \mathcal{N}$. The set of time steps is defined as $\mathcal{T} = \{0, \ldots, T\}$ with index t. Without loss of generality, every bus is assumed to have a load demand, a energy storage and a renewable source. To obtain a polynomial formulation, bus voltages are expressed in rectangular coordinates. Let e_{kt} and f_{kt} denote real and imaginary parts of complex bus voltage at bus k at time step t. Further define $\boldsymbol{v}_{kt} = (e_{kt}, f_{kt})$ and $\boldsymbol{v}_t = (\boldsymbol{v}_{kt})_{k\in\mathcal{N}}$. The $(l,m)^{th}$ element of the network admittance matrix is denoted by $g_{lm} + \mathrm{j}b_{lm}$. To state the power balance equation, we define:

$$\begin{aligned} p_k(\boldsymbol{v}_t) = \sum_{(j,k)\in\mathcal{L}} &\{(g_{jk}e_{it} + b_{jk}f_{it})(e_{jt} - e_{kt}) \\ &+(g_{jk}f_{jt} - b_{jk}e_{jt})(f_{jt} - f_{kt})\} + p^d_{kt} - p^r_{kt} \end{aligned} \tag{7-1}$$

$$\begin{aligned} q_k(\boldsymbol{v}_t) = \sum_{(j,k)\in\mathcal{L}} &\{(g_{jk}f_{jt} - b_{jk}e_{jt})(e_{jt} - e_{kt}) \\ &-(g_{jk}e_{jt} + b_{jk}f_{jt})(f_{jt} - f_{kt})\} + q^d_{kt} - q^r_{kt} \end{aligned} \tag{7-2}$$

where p^d_{kt} and q^d_{kt} are the active and reactive power of load demand at bus k and time t; p^r_{kt} and q^r_{kt} are the active and reactive power of renewable energy generation at bus k and time t. To state the line flow constraint, we define

$$\begin{aligned} i_{jk}^2(\boldsymbol{v}_t) = &\left(g_{jk}(e_{jt}-e_{kt}) - b_{jk}(f_{jt}-f_{kt})\right)^2 \\ &+ \left(b_{jk}(e_{jt}-e_{kt}) - g_{jk}(f_{jt}-f_{kt})\right)^2. \end{aligned} \tag{7-3}$$

Based on the definitions above, the multi-period OPF problem is formally stated as

$$\min_{\Omega} \sum_{t\in\mathcal{T}}\sum_{k\in\mathcal{G}} f_{kt}^g \quad \text{s.t.} \tag{7-4a}$$

$$p_{kt}^g = p_k(\boldsymbol{v}_t) - r_{kt}^d + r_{kt}^c,\ \forall k \in \mathcal{N} \tag{7-4b}$$

$$q_{kt}^g = q_k(\boldsymbol{v}_t),\ \forall k \in \mathcal{N} \tag{7-4c}$$

$$\underline{P}_k^g \le p_{kt}^g \le \overline{P}_k^g,\ \forall k \in \mathcal{N} \tag{7-4d}$$

$$\underline{Q}_k^g \le q_{kt}^g \le \overline{Q}_k^g,\ \forall k \in \mathcal{N} \tag{7-4e}$$

$$\underline{V}_k^2 \le e_{kt}^2 + f_{kt}^2 \le \overline{V}_k^2,\ \forall k \in \mathcal{N} \tag{7-4f}$$

$$i_{jk}^2(\boldsymbol{v},t) \le \overline{I}_{jk}^2,\ \forall (j,k) \in \mathcal{L} \tag{7-4g}$$

$$s_{kt} = s_{k(t-1)} + \left(\eta_k^c r_{kt}^c - r_{kt}^d/\eta_k^d\right)\Delta t,\ \forall k \in \mathcal{N} \tag{7-4h}$$

$$0 \le r_{kt}^c \le R_k^c,\ \forall k \in \mathcal{N} \tag{7-4i}$$

$$0 \le r_{kt}^d \le R_k^d,\ \forall k \in \mathcal{N} \tag{7-4j}$$

$$0 \le s_{kt} \le S_k,\ \forall k \in \mathcal{N} \tag{7-4k}$$

$$s_{k0} = s_{kT} = S_k^0,\ \forall k \in \mathcal{N} \tag{7-4l}$$

$$-RR_k \le p_{kt}^g - p_{k(t-1)}^g \le RR_k,\ \forall k \in \mathcal{G} \tag{7-4m}$$

$$r_{kt}^d r_{kt}^c = 0,\ \forall k \in \mathcal{N} \tag{7-4n}$$

where $\Omega = \left(\boldsymbol{v}_{kt}, s_{kt}, r_k^{dt}, r_{kt}^c, p_{kt}^g, q_{kt}^g\right)_{k\in\mathcal{N},t\in\mathcal{T}}$ denotes the decision variable space and $f_{kt}^g = c_{k2}(p_{kt}^g)^2 + c_{k1}p_{kt}^g + c_{k0}$ represents the operation cost of conventional generator at bus k and time t. (7-4b) and (7-4c) are the power balance equations with p_{kt}^g and q_{kt}^g denoting the active and reactive power output of conventional generator at bus k and time t. The upper and lower bounds of generator active and reactive power output are enforced by (7-4d) and (7-4e). If no conventional generator is installed at bus t, both the upper and lower bounds in (7-4d) and (7-4e) are set to zeros. (7-4f) is the bus voltage magnitude constraint and (7-4g) is the line flow constraint. The storage dynamics are given by (7-4h) where η_k^c and η_k^d denote the charge and discharge efficiency; (7-4i), (7-4j) and (7-4k) set the limits for charge power, discharge power and energy storage level of each ESS, respectively. (7-4l) provides the boundary condition for ESS by setting the initial and final stored energy to a pre-specified common value. (7-4m) represents the conventional generator ramping constraint, i.e. the output active power change between two consecutive time steps should be within certain limit. In a congested power network, the locational marginal price (LMP) at some buses can be negative which makes simultaneously charging and discharging of ESSs at those buses a more economical operational situation. But this is unrealistic for most storage technologies. To avoid this, we

add equality constraint (7-4n) to ensure ESSs operate either in the charge or discharge mode.

To simplify notation in the discussion followed, we rewrite problem (7-4) in a compact form as

$$\min_{\Omega} \sum_{t\in\mathcal{T}} f(\boldsymbol{v}_t) \quad \text{s.t.} \tag{7-5a}$$

$$\boldsymbol{g}(\boldsymbol{v}_t, \boldsymbol{u}_t) \geq \boldsymbol{0},\ \forall t \in \mathcal{T} \tag{7-5b}$$

$$(\boldsymbol{u}_t)_{t\in\mathcal{T}} \in \mathbb{K} \tag{7-5c}$$

where $\boldsymbol{u}(t) = \left(s_{kt}, r^d_{kt}, r^c_{kt}, p^g_{kt}, q^g_{kt}\right)_{k\in\mathcal{N}}$ and $f(\boldsymbol{v}_t) = \sum_{k\in\mathcal{G}} f^g_k(\boldsymbol{v}_t)$. By observing a equality constraint can be denoted by two inequality constraints, (7-5b) represents all constraints in problem (7-4) except for (7-4h) and (7-4m). Linear equality (7-4h)(7-4l) and inequality (7-4m) together define a polyhedron, a simple convex set, denoted as $\mathbb{K}$ in (7-5c). Note that the objective function (7-5a) and constraint (7-5b) are decoupled in time, while only constraint (7-5c) is coupled in time.

7.3 Parallel Moment Approach to Multi-Period OPF Problem

7.3.1 ADMM Decomposition

By adding auxiliary variable $(\boldsymbol{z}(t))_{t\in\mathcal{T}}$, problem (7-5) is equivalently transformed into

$$\min_{\Omega} \sum_{t\in\mathcal{T}} f(\boldsymbol{v}_t) \quad \text{s.t.} \tag{7-6a}$$

$$\boldsymbol{g}(\boldsymbol{v}_t, \boldsymbol{u}_t) \geq \boldsymbol{0},\ \forall t \in \mathcal{T} \tag{7-6b}$$

$$(\boldsymbol{u}_t)_{t\in\mathcal{T}} = (\boldsymbol{z}_t)_{t\in\mathcal{T}} \tag{7-6c}$$

$$(\boldsymbol{z}_t)_{t\in\mathcal{T}} \in \mathbb{K}. \tag{7-6d}$$

To further simplify notation, we denote $\boldsymbol{v} = (\boldsymbol{v}_t)_{t\in\mathcal{T}}$, and $\boldsymbol{u}$, $\boldsymbol{z}$ and $\boldsymbol{\lambda}$ are defined in similar way. The augmented Lagrangian function related to equality constraint (7-6c) thus is given by

$$\begin{aligned} &L\left(\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{z}, \boldsymbol{\lambda}\right) \\ =&\sum_{t\in\mathcal{T}} f(\boldsymbol{v}_t) + \boldsymbol{\lambda}^T\left(\boldsymbol{u} - \boldsymbol{z}\right) + \frac{\rho}{2}\|\boldsymbol{u} - \boldsymbol{z}\|_2^2. \end{aligned} \tag{7-7}$$

According to the alternating direction method of multipliers (ADMM) [136], we can sovle problem (7-6) by the following iterations:

$(\boldsymbol{v}, \boldsymbol{u})$-subproblem

$$\left(\boldsymbol{v}^{k+1}, \boldsymbol{u}^{k+1}\right) := arg\min_{(\boldsymbol{v},\boldsymbol{u})} L\left(\boldsymbol{v}, \boldsymbol{u}, \boldsymbol{z}^k, \boldsymbol{\lambda}^k\right) \text{ s.t. } (7-6b) \tag{7-8}$$

$\boldsymbol{z}$-subproblem

$$\boldsymbol{z}^{l+1} := arg\min_{\boldsymbol{z}} L\left(\boldsymbol{v}^{k+1}, \boldsymbol{u}^{k+1}, \boldsymbol{z}, \boldsymbol{\lambda}^{k}\right) \text{ s.t. } (7-6d) \tag{7-9}$$

$\boldsymbol{\lambda}$-update

$$\boldsymbol{\lambda}^{k+1} := \boldsymbol{\lambda}^{k} + \rho\left(\boldsymbol{u}^{k+1} - \boldsymbol{z}^{k+1}\right) \tag{7-10}$$

ADMM is proved to converge to the global minimum of problem (7-6) under the convexity of problem (7-8) and problem (7-9). In the current formulation, problem (7-9) is obviously convex due to the convex quadratic objective function and the polyhedron feasible set. But problem (7-8) is still a non-convex problem, we will relax it to a convex problem in the next subsection. The convergence of ADMM can be measured by

$$\Gamma^{l} = \left\| \begin{matrix} \boldsymbol{z}^{k+1} - \boldsymbol{z}^{k} \\ \boldsymbol{x}^{k+1} - \boldsymbol{z}^{k+1} \end{matrix} \right\|. \tag{7-11}$$

The ADMM iteration terminates when $\Gamma^{l} < \varepsilon$.

7.3.2 Parallel Moment Relaxation of (v,x)-Subproblem

Problem (7-8) is separable among different time steps thus can be solved in parallel. The problem associated with time t is given by

$$\begin{aligned} &\min_{\boldsymbol{v}_t, \boldsymbol{u}(t)} h^{k}(\boldsymbol{v}_t, \boldsymbol{u}_t) \\ &\text{s.t.} \quad \boldsymbol{g}(\boldsymbol{v}_t, \boldsymbol{u}_t) \geq \boldsymbol{0}. \end{aligned} \tag{7-12}$$

where $h^{k}(\boldsymbol{v}_t, \boldsymbol{u}_t) = f(\boldsymbol{v}_t) + \left(\boldsymbol{\lambda}_t^{k}\right)^{T}\left(\boldsymbol{u}_t - \boldsymbol{z}(t)^{k}\right) + \frac{\rho}{2}\|\boldsymbol{u}_t - \boldsymbol{z}_t^{k}\|_2^2$. Problem (7-12) is a nonlinear, non-convex polynomial optimization problem. Due to development in polynomial optimization and real algebraic geometry, a moment-sos approach has been proposed in literature to construct a hierarchy of semidefinite program (SDP) whose optima asymptotically and often finitely converge to the global optimum of the original non-convex problem. Recent study has revealed the applicability and exactness of moment approach to single period OPF problem. For convenience, we define $\boldsymbol{x}(t) = (\boldsymbol{v}_t, \boldsymbol{u}_t)$ and drop the time index t to mean the statement holds for arbitrary $t \in \mathcal{T}$ in the following discussion in Section-III-B and Section-III-C. Thus problem (7-12) is simply denoted as $\min_{\boldsymbol{x}} h^{k}(\boldsymbol{x})$ s.t. $\boldsymbol{g}(\boldsymbol{x}) \geq \boldsymbol{0}$.

To formulate the moment relaxation of problem (7-12), we need the following definitions. Every monomial of the decision variable $\boldsymbol{x}$, denoted as $\boldsymbol{x}^{\boldsymbol{\alpha}}$, can be associated with a moment variable $y(\boldsymbol{\alpha})$ through certain Riesz linear functional $L_{\boldsymbol{y}} : \mathbb{R}[\boldsymbol{x}] \mapsto \mathbb{R}$, i.e. $L_{\boldsymbol{y}}(\boldsymbol{x}^{\boldsymbol{\alpha}}) = y(\boldsymbol{\alpha})$ where $\boldsymbol{y}$ denotes the infinite dimensional moment sequence $(y(\boldsymbol{\alpha}))_{\boldsymbol{\alpha}\in\mathbb{N}^n}$. The moment matrix $M_d(\boldsymbol{y})$ associated with $\boldsymbol{y}$ is the real symmetric matrix with rows and columns indexed in certain monomial basis $(\boldsymbol{x}^{\boldsymbol{\alpha}})$ with entries $M_d(\boldsymbol{y})(\boldsymbol{\alpha}, \boldsymbol{\beta}) = L_{\boldsymbol{y}}(\boldsymbol{x}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}) = y(\boldsymbol{\alpha}+\boldsymbol{\beta})$, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{N}_d^n$.

Similarly, the localizing matrix $M_d(g\boldsymbol{y})$ associated with $\boldsymbol{y}$ and polynomial $g \in \mathbb{R}[\boldsymbol{v}]$ is the real symmetric matrix with rows and columns indexed in monomial basis $(\boldsymbol{x}^{\boldsymbol{\alpha}})$ with entries $M_d(g\boldsymbol{y})(\boldsymbol{\alpha},\boldsymbol{\beta}) = L_{\boldsymbol{y}}(g(\boldsymbol{x})\boldsymbol{v}^{\boldsymbol{\alpha}+\boldsymbol{\beta}}) = \sum_{\boldsymbol{\gamma}} g_{\boldsymbol{\gamma}} y(\boldsymbol{\alpha}+\boldsymbol{\beta}+\boldsymbol{\gamma})$, $\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{N}^n_d$. Under these definitions, the order-d moment relaxation of problem (7-12) is stated as

$$\min_{\boldsymbol{y}\in\mathbb{R}^{\binom{n+2d}{2d}}} L_{\boldsymbol{y}}(h^k) \quad \text{s.t.} \tag{7-13a}$$

$$M_d(\boldsymbol{y}) \succcurlyeq \mathbf{0} \tag{7-13b}$$

$$M_{d-1}(g\boldsymbol{y}) \succcurlyeq \mathbf{0} \tag{7-13c}$$

$$y(\mathbf{0}) = 1 \tag{7-13d}$$

Note that polynomial optimization problem (7-12) only involves monomials with order less or equal to two (see details in (7-4)), therefore order-1 moment relaxation is enough to cover all the monomials. After solving the SDP (7-13) to obtain $\boldsymbol{y}^*$, the solution of the original problem (7-12) can be extracted from the spectral decomposition of the diagonal block of $M_d(\boldsymbol{y}^*)$ related to the second-order terms, i.e. $\boldsymbol{v}^* = \sqrt{\lambda_1}\boldsymbol{\eta}_1$ where λ_1 is the largest eigenvalue with related eigenvector $\boldsymbol{\eta}_1$ of $L_{\boldsymbol{y}^*}(\boldsymbol{x}\boldsymbol{x}^T)$. When moment matrix $M_d(\boldsymbol{y}^*)$ has rank-1, $\boldsymbol{v}^*$ is the theoretically guaranteed global solutioin of (7-12). Even if rank-1 condition is not strictly satisfied, $\boldsymbol{v}^*$ is an approximate solution to (7-12).

7.3.3 Exploiting Sparsity in (v,x)-Subproblem

The dimensions of decision variables and semidefinite constraints in problem (7-13) increase drastically as the system scale and relaxation order increase and quickly become prohibitively high for current SDP solvers. To improve the numerical tractability of moment relaxation (7-13), the sparsity in problem (7-12) must be exploited to reduce problem size.

The sparsity in problem (7-12) is described by a $n \times n$ symmetric symbolic matrix $\boldsymbol{R}$. Its element R_{ij} is nonzero value $\star$ if and only if either (i) $i = j$, or (ii) x_i and x_j appear simultaneously in a monomial of $h^k(\boldsymbol{x})$, or (iii) x_i and x_j appear simultaneously in a polynomial component of $\boldsymbol{g}(\boldsymbol{x})$. All other element of $\boldsymbol{R}$ are zeros. The graph which takes $\boldsymbol{R}$ as its adjacency matrix is called the correlative sparsity pattern (CSP) graph of polynomial optimization problem (7-12). The following strategy is proposed in [172] to partition the decision variables into several possibly overlapping groups to significantly reduce the problem size of moment relaxation while keep its asymptotical convergence property. First, generate a chordal extension of the CSP graph. Then find all maximal cliques of the chordal extension which form a required partition of the decision variables.

Considering the special structure of the problem discussed in this chapter, we propose the following procedure to generate the required variable partition. Step 1, consider the CSP subgraph $G(\mathcal{V},\mathcal{E})$ related to the subset of decision variable $\mathcal{V} = (\boldsymbol{v}_i)_{i\in\mathcal{N}}$. There is a edge

between $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ if either bus i and bus j are connected or they both connect to a third bus. We denote the associated adjacency matrix as $\boldsymbol{Q}$. Let $G(\mathcal{V}, \overline{\mathcal{E}})$ be the chordal extension of $G(\mathcal{V}, \mathcal{E})$. The maximal cliques of chordal graph $G(\mathcal{V}, \overline{\mathcal{E}})$ can be determined through Cholesky factorization of $\boldsymbol{Q} + \delta \boldsymbol{I}$ under the perfect elimination ordering where δ is a positive number to guarantee $\boldsymbol{Q} + \delta \boldsymbol{I}$ is positive definite. We denote the maximal cliques of $G(\mathcal{V}, \overline{\mathcal{E}})$ as sets $\mathcal{V}_1, \ldots, \mathcal{V}_p$. Step 2, consider $\boldsymbol{u}_k = \left(s_k, r_k^d, r_k^c, p_k^g, q_k^g\right)$ for any $i \in \mathcal{N}$. Observe that $\mathcal{W}_k$, r_k^d, r_k^c and p_k^g connect to each other through (7-4b) where $\mathcal{W}_k = \{\boldsymbol{v}_j | j = k \text{ or bus } j \text{ connect to bus } k\}$. By definition, $\mathcal{W}_k$ is a clique in $G(\mathcal{V}, \mathcal{E})$ thus there exist $\mathcal{V}_l$ such that $\mathcal{W}_k \subseteq \mathcal{V}_i$. In addition, q_k^g also connects to $\mathcal{W}_k$ through (7-4c). We visualize the connections among $\mathcal{W}_k$, r_k^d, r_k^c, p_k^g and q_k^g in Fig. 1 where solid lines represent the connection described above and dashed lines are connections need to be added to obtained a complete graph. Combine the results of step 1 and

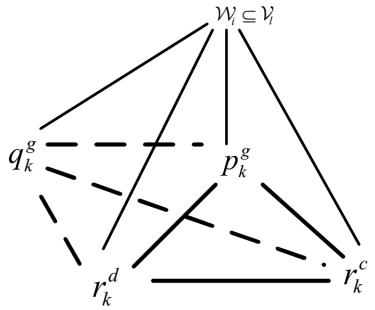


Figure 7-1 Clique Structure in $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$

step 2, we obtain the required variable partition as $\mathcal{X}_i = \mathcal{V}_i \bigcup \left(\bigcup_{k:\mathcal{W}_k \subseteq \mathcal{V}_i} \boldsymbol{u}_k\right), i = 1, \ldots, p$.

Then the sparsity-exploiting moment relaxation of polynomial optimization problem is given by

$$\min_{\boldsymbol{y}} \text{vec}(h^k)^T \boldsymbol{y} \quad \text{s.t.} \tag{7-14a}$$

$$M_d(\boldsymbol{y}, \mathcal{X}_i) \succcurlyeq \boldsymbol{0}, \quad \forall i = 1, \ldots, p \tag{7-14b}$$

$$M_{d-1}(g\boldsymbol{y}, \mathcal{X}_i) \succcurlyeq \boldsymbol{0}, \quad \forall i = 1, \ldots, p \tag{7-14c}$$

$$y(\boldsymbol{0}) = 1 \tag{7-14d}$$

where $M_d(\boldsymbol{y}, \mathcal{X}_i)$ (resp. $M_{d-d_g}(g\boldsymbol{y}, \mathcal{X}_i)$) is the moment (resp. localizing) submatrix obtained from $M_d(\boldsymbol{y})$ (resp. $M_{d-d_g}(g\boldsymbol{y})$) by retaining those rows and columns indexed by monomials only involve variables in $\mathcal{X}_i$. The sparse moment relaxation is exact if rank $M_{d_i}(\boldsymbol{y}, \mathcal{X}_i) = 1, \forall 1 \leq i \leq p$. The solution x_k can then be extracted from the spectral decomposition of the diagonal block corresponding to second-order monomials in $M_{d_i}(\boldsymbol{y}, \mathcal{X}_i)$ where $x_k \in \mathcal{X}_i$.

7.3.4 Solution of z-Subproblem

Observe that the $\boldsymbol{z}$-subproblem (7-9) is separable in space, thus the optimization can be performed in parallel among different buses. For convenience, we drop the index k to mean

the statement holds for every bus in this subsection. Considering the definition of augmented Lagrangian function (7-7) and polyhedron $\mathbb{K}$, using perfect square formula, $\boldsymbol{z}$-subproblem (7-9) can be equivalently written as

$$\min_{\boldsymbol{z}} \|\boldsymbol{z} - \left(\boldsymbol{u}^{k+1} + \frac{1}{\rho}\boldsymbol{\lambda}^k\right)\|_2^2 \quad \text{s.t. } \boldsymbol{z} \in \mathbb{K} \tag{7-15}$$

Problem (7-15) is actually a projection of $\boldsymbol{u}^{k+1} + \frac{1}{\rho}\boldsymbol{\lambda}^k$ onto the polyhedron $\mathbb{K}$. Although no analytical solution exists in general, it can be solved effectively with standard IPM solver.

7.4 Case Study

In this section, we report a case study on IEEE 14-bus system with 3 energy storages and 2 wind farms. Energy storages are located at bus 4, 5 and 7, and wind farms are installed at bus 2 and 3. The diagram of the system is shown in Fig. 7-2. The energy capacity for each storage is 32 MW, and the charge & discharge power capacity are both set to be 8 MW/h. The charge & discharge efficiency are both 0.9. Each wind farm has a maximal output power of 40 MW. We consider a time horizon of 24 h with each time step 30 min. The load and wind generation generation follow the forecast curve shown in Fig. 7-3. The objective is minimizing the conventional generation cost.

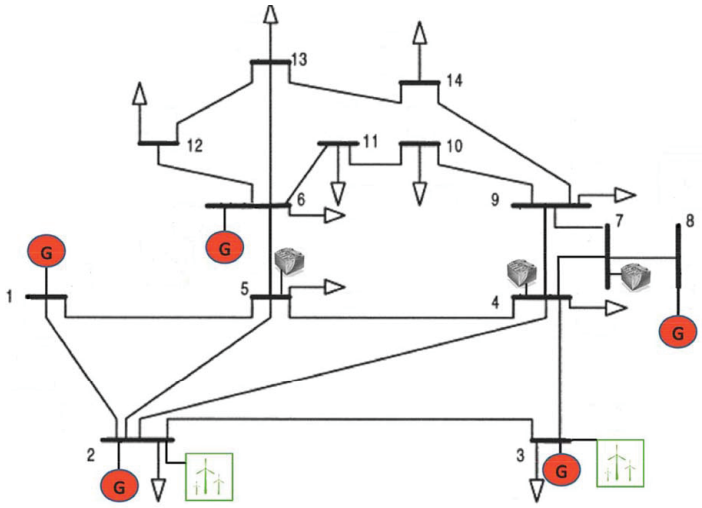


Figure 7-2 IEEE 14-bus System with ESSs and Wind Generation

The algorithm is programmed in MATLAB with YALMIP as the modeling tool and SDPT3 as SDP and QP solver, running on a Win8 PC with a 3.0 GHz CPU with 8GB RAM. In this case, first-order moment relaxation is applied to each $(\boldsymbol{v}, \boldsymbol{u})$-subproblem, and the satisfaction of rank-1 condition is observed through the solution process. The total dimension of the decision variables is 451×48 and the total total size of the linear matrix inequality (LMI) constraints is 102×48, which is intractable for state-of-art SDP solvers if no decomposition techniques are implemented. It is noteworthy that if smaller time period is used, e.g. 5 min leading to

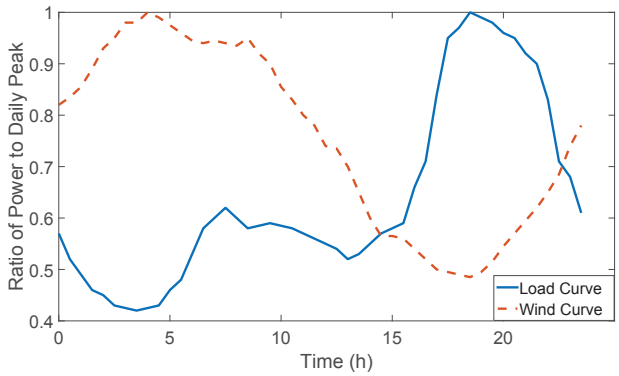


Figure 7-3 Load and Wind Generation Forecast

288 steps as considered in [91], the size of LMI will be even more hopelessly intractable for naive application of any SDP solver. However, the problem is still tractable if the decomposition algorithm framework discussed in this chapter is implemented since the size of the $(\boldsymbol{v}, \boldsymbol{u})$-subproblem is irrevelant to the number of time steps.

Fig. 7-4 shows the decreasing process of convergence measure (7-11). It takes 106 iterations to reach the tolerance $\varepsilon = 10^{-5}$. In this preliminary test, the algorithm, though parallel in nature, is only implemented sequentially. Under such condition, the computation time required is 2.03 h. Fig. 7-5 shows the optimal operational strategies for all ESSs and generators. It shows that one major charge & discharge cycle and two minor cycles are needed to achieve optimal operation. In the major cycle, ESSs charge with maximal power during 2:00 to 5:00 and discharge with maximal power during 17:00 to 22:00. This major cycle basically coincides with the load and wind generation variation pattern shown in Fig. 7-3. Under the given load level and wind generation, only three out of five generators have been started up during the operation period, which indicates the priority of wind power utilization in current problem settings.

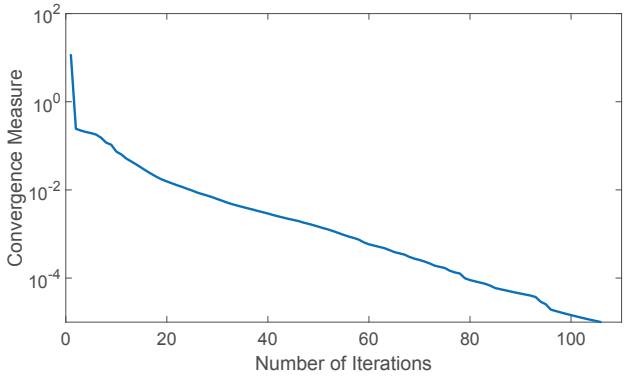
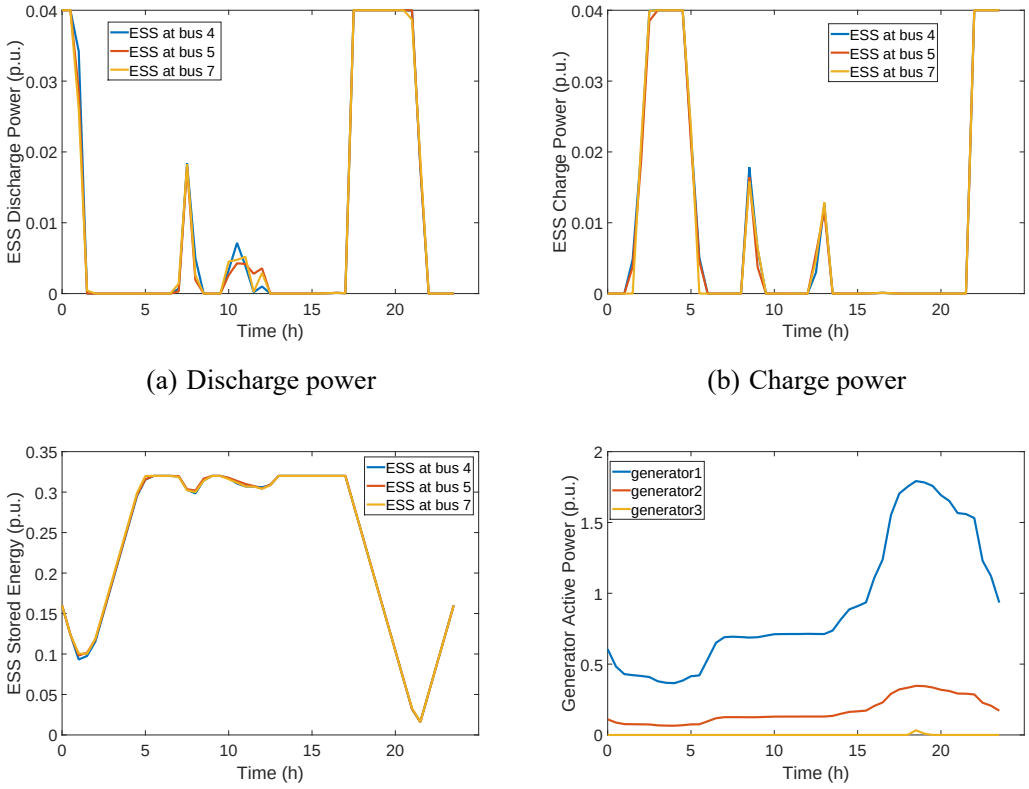


Figure 7-4 IEEE 14-bus System with ESSs and Wind Generation



(a) Discharge power (b) Charge power

(c) Stored energy (d) Generator output

Figure 7-5 Optimal Operation Curve of ESSs and Generators

7.5 Conclusions

In this chapter, a decomposition algorithm have been designed to solve full AC multi-period OPF problem with energy storage devices and renewable generations, based on moment relaxation and ADMM. Case study on IEEE 14-bus system shows that the decomposition algorithm can effectively solve the otherwise intractable problem. Future work includes parallel implementation of the algorithm in high performance computing clusters and application of warm-start strategy [181] to IPM solver for subproblems and using fast ADMM iterations [182] to accelerate convergence. Incorporating uncertainties into the proposed framework would be another interesting direction for future research.

8 Structure-exploiting Delay-dependent Stability Analysis of Power System Load Frequency Control

8.1 Introduction

The existence and adverse effects of time delays in power system control loops have been well-recognised by recent publications with the ever-increasing integration of communication networks into the closed-loop control & operation of the currently developing smart grids. Time delays can degrade the dynamic performance and stability of load frequency control (LFC) [57, 58, 60, 62, 183]. To analyze and further alleviate the effects of time delays, many researchers have focused their attention on delay margin calculation and time-delay robust control.

Delay margins are the maximal admissible time delays with which the system remains stable [57, 58]. The knowledge of the delay margins can be used to evaluate the stability level and guide the controller design of time-delay power systems. Two groups of methods, including frequency-domain and time-domain methods, have been proposed to calculate the delay margins. Frequency-domain methods obtain the delay margins by computing all critical roots of system characteristic equations [184, 185]. This group of methods are also successfully applied to time-delay power systems [183, 186–190]. Although frequency-domain methods can obtain exact delay margins, they are limited to constant delays which rarely occur in practice. Time-domain methods establish sufficient conditions for the stability of time-delay systems by constructing Lyapunov-Krasovskii functionals (LKF) whose parameters can be determined by solving linear matrix inequalities (LMI) [59]. This group of methods are applied to LFC with constant and time-varying delays [57, 58]. LMI-based time domain methods can even be extended to analyze the stability of nonlinearly perturbed LFC with time delays [191]. Although time-domain methods possess some conservatism when calculating the delay margins, they can be conveniently applied to systems with single or multiple, constant or time-varying delays. Moreover, the development of the modern interior point method (IPM) provides off-the-shelf tools to solve LMIs.

The ultimate goal of DDSA is to inform the controller design to mitigate the adverse effects of time delays. From this aspect, time-domain LMI based methods have advantages over frequency domain methods due to the fact that LMI based criteria can be conveniently employed to derive the delay-independent and delay-dependent bounded real lemmas (BRL) [59] which then can be used to design H_∞ robust controllers. For instance, delay-independent BRLs are used in [60] and [61] to design state-feedback and PI robust controllers for LFC, respectively. In [62], PID-type robust controllers are designed for LFC systems based on a delay-dependent BRL. The delay-dependent BRL in [192] is extended to consider LFC pa-

rameter uncertainties. To reduce conservatism, some probabilistic information of time delays are taken into account in the delay-dependent BRL for robust LFC design in [193]. Above-mentioned controllers are all designed to minimize the H_∞ index while guarantee stability for any delays less than the preset upper bounds (the upper bounds are ∞ for delay-independent methods).

However, there is still a long way to go before all those analysis/synthesis methods for time-delay systems can be applied to real-world power systems. The major obstacle for time-domain LMI based methods is the computational burden of solving large-scale LMIs. Due to the limited ability of the state-of-art LMI solvers, LMI based methods are only applied to power systems with less than 3 generation units in each control area [57, 58, 62] or using order-reduced system models [56**?**]. As we know, there could be hundreds of generation units in a practical power system and the analysis based on order-reduced system models may not obtain reliable results. Therefore, this chapter focuses on the computational aspect and aims at improving the numerical tractability of LMI based DDSA. All previous works in this field directly apply certain methods to LFC without considering the specific structure of power system control loops. Nevertheless, problems appear in the power systems are not generic ones but rather present some structures. Thus, if those structures are properly exploited, more efficient and feasible computational methods can be obtained.

The contribution of this chapter is exploiting the structures, including chordal sparsity and symmetry, of LFC loops to significantly improve the numerical tractability of DDSA. The idea of exploiting chordal sparsity in Lyapunov LMI has already been discussed in [194]. We further extend this idea to DDSA by proving some more general results concerning the structure restrictions of weighting matrices needed for the LMIs in DDSA to inherit the chordal sparsity of LFC loops. Chordal sparsity in the LMIs then allows the decomposition of the original positive semi-definite (PSD) condition into much smaller ones. Moreover, the symmetry of LFC control loops is exploited to reduce the number of decision variables. By exploiting chordal sparsity and symmetry, the number of decision variables and size of PSD conditions are greatly reduced. Numerical results show the proposed structure-exploiting techniques significantly improves the numerical tractability of DDSA applied to large-scale LFC problems at the price of introducing minor and acceptable conservatism.

The rest of this chapter is organized as follows. Section 8.2 describes the structure of time-delay LFC systems and the delay-dependent stability criterion. In section 8.3, we present the chordal-structure matrix decomposition (CSMD) and give several lemmas which enable the application of CSMD to DDSA. Section 8.4 introduces the sparsity and symmetry exploiting techniques in DDSA of LFC. Step-by-step implementation guideline is also given in this section. Section 8.5 reports numerical studies. Possible extensions of the proposed method is discussed in section 8.6. Finally, section 8.7 draws conclusions and gives suggestions on

future research.

8.2 Time-delay Load Frequency Control and Delay-dependent Stability Analysis

8.2.1 Load Frequency Control Loops

As the main function of automatic generation control (AGC) system, LFC aims at maintaining frequency and power interchanges between neighboring control areas at scheduled values [61]. Both the single-area and multi-area control schemes are shown in Fig. 8-1. The modeling of LFC system is mainly based on two simplifications [61, 111]: 1) LFC focuses on the collective performance of all generators rather than the inter-machine oscillations, therefore all generators in one control area are aggregated into one generator with inertia M_i equal to the sum of the inertia of all generators and the damping effect of all loads is represented by a single damping constant D_i; 2) due to the first simplification and the slow dynamics of prime-mover power change and system frequency response, the dynamics of electrical angle voltages are all neglected. As a result, the LFC model consists of governor, turbine and rotor/load, and a PI controller is implemented in practice to guarantee system stability and desirable dynamic performance. The time delays, including communication delays, sample-induced delays and fault-induced delays, are combined as one delay in the control loop represented by the block $e^{-s\tau_i}$. The formulation of the state-space model can be found in appendix A. The control loops are naturally associated with a graph representing the connection relationship among state variables, shown in Fig. 8-2. In this chapter, we will exploit the structure characteristics of this graph, including chordal sparsity and symmetry, to improve the numerical tractability of DDSA of LFC.

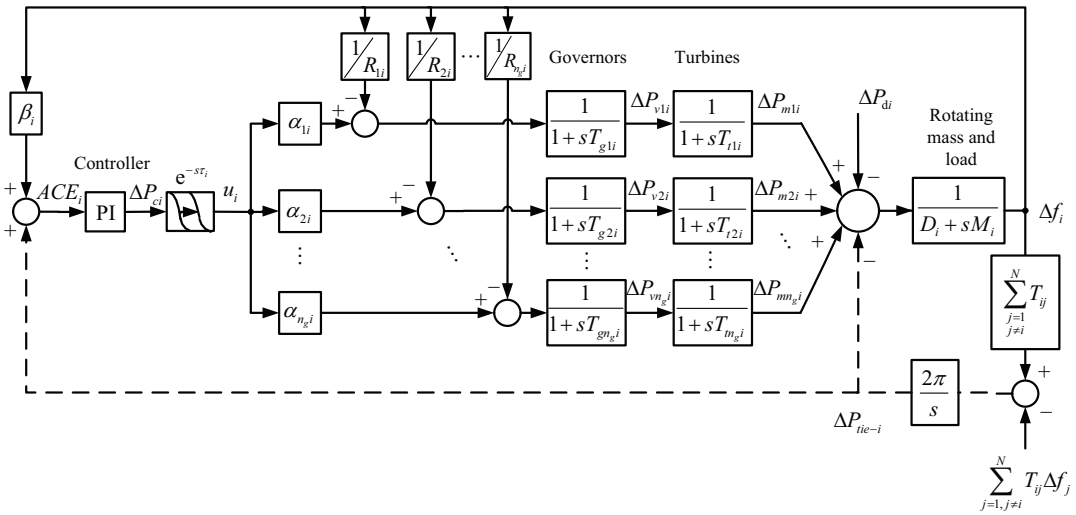


Figure 8-1 Single-area (without dotted lines) and Multi-area (with dotted lines) LFC control structure.

The state-space model for the open-loop system in area i is

$$\begin{cases} \dot{x}_i(t) &= \boldsymbol{A}_i x_i(t) + \sum_{j=1,j\neq i}^{N} \boldsymbol{A}_{ij} x_j(t) + \boldsymbol{B}_i u_i(t-\tau_i) \\ y_i(t) &= \boldsymbol{C}_i x_i(t) \end{cases} \tag{8-1}$$

where

$$x_i^\top = \left[\Delta f_i, \Delta P_{tie-i}, \Delta P_{m1i}, \cdots, \Delta P_{mn_g i}, \Delta P_{v1i}, \cdots, \Delta P_{vn_g i}\right]$$

$$y_i = ACE_i$$

$$\boldsymbol{A}_i = \begin{bmatrix} \boldsymbol{A}_{11i} & \boldsymbol{A}_{12i} & \boldsymbol{0}_{2\times n_g} \\ \boldsymbol{0}_{n_g\times 2} & \boldsymbol{A}_{22i} & \boldsymbol{A}_{23i} \\ \boldsymbol{A}_{31i} & \boldsymbol{0}_{n_g\times n_g} & \boldsymbol{A}_{33i} \end{bmatrix}$$

$$\boldsymbol{A}_{ij} = \begin{bmatrix} 0 & 0 & \boldsymbol{0}_{1\times 2n_g} \\ -2\pi T_{ij} & 0 & \boldsymbol{0}_{1\times 2n_g} \\ \boldsymbol{0}_{2n_g\times 1} & \boldsymbol{0}_{2n_g\times 1} & \boldsymbol{0}_{2n_g\times 2n_g} \end{bmatrix}$$

$$\boldsymbol{B}_i = \begin{bmatrix} \boldsymbol{0}_{2\times 1} \\ \boldsymbol{0}_{n_g\times 1} \\ \boldsymbol{B}_{3i} \end{bmatrix}, F_i = \begin{bmatrix} \frac{-1}{M_i} \\ \boldsymbol{0}_{(1+n_g)\times 1} \\ \boldsymbol{0}_{n_g\times 1} \end{bmatrix}$$

$$\boldsymbol{C}_i = \begin{bmatrix} \beta_i & 1 & \boldsymbol{0}_{1\times 2n_g} \end{bmatrix}, D_i = \begin{bmatrix} 0 & -1 & \boldsymbol{0}_{1\times n_g} \end{bmatrix}$$

$$\boldsymbol{A}_{11i} = \begin{bmatrix} -\frac{D_i}{M_i} & -\frac{1}{M_i} \\ 2\pi \sum_{j=1,j\neq i}^{N} T_{ij} & 0 \end{bmatrix}, \boldsymbol{A}_{12i} = \begin{bmatrix} \frac{1}{M_i} & \cdots & \frac{1}{M_i} \\ 0 & \cdots & 0 \end{bmatrix}$$

$$\boldsymbol{A}_{22i} = -\boldsymbol{A}_{23i} = diag\left\{ -\frac{1}{T_{t1i}}, \cdots, -\frac{1}{T_{tn_g i,}} \right\}$$

$$\boldsymbol{A}_{31i} = \begin{bmatrix} \frac{-1}{T_{g1i}R_{1i}} & \cdots & \frac{-1}{T_{gn_g i}R_{n_g i}} \\ 0 & \cdots & 0 \end{bmatrix}^\top \boldsymbol{B}_{3i} = \begin{bmatrix} \frac{\alpha_{1i}}{T_{g1i}} & \cdots & \frac{\alpha_{n_g i}}{T_{gn_g i}} \end{bmatrix}^\top$$

$$\boldsymbol{A}_{33i} = diag\left\{ \frac{-1}{T_{g1i}}, \cdots, \frac{-1}{T_{gn_g i}} \right\}$$

The PI controller in area i takes the form

$$u_i(t) = -K_{Pi} ACE_i - K_{Ii} \int ACE_i dt \tag{8-2}$$

To simplify the analysis, further define the virtual state vectors $\bar{x}_i = [x_i^\top, \int y_i^\top]^\top$, the closed-loop system can then be rewritten as

$$\dot{\bar{x}}_i(t) = \bar{\boldsymbol{A}}_{ii}\bar{x}_i(t) + \bar{\boldsymbol{A}}_{dii}\bar{x}_i(t-\tau_i) + \sum_{j=i,j\neq i}^{N} \bar{\boldsymbol{A}}_{ij}\bar{x}_j(t) \tag{8-3}$$

where

$$\bar{\boldsymbol{A}}_{ii} = \begin{bmatrix} \boldsymbol{A}_i & 0 \\ \boldsymbol{C}_i & 0 \end{bmatrix}, \ \bar{\boldsymbol{A}}_{dii} = \begin{bmatrix} -K_{Pi}\boldsymbol{B}_i\boldsymbol{C}_i & -K_{Ii}\boldsymbol{B}_i \\ \boldsymbol{0}_{2n_g+2} & 0 \end{bmatrix}$$

$$\bar{\boldsymbol{A}}_{ij} = \begin{bmatrix} \boldsymbol{A}_{ij} & 0 \\ 0 & 0 \end{bmatrix}$$

By defining the state vector as $x = [\bar{x}_1^\top, \bar{x}_2^\top, \cdots, \bar{x}_n^\top]^\top$, model (8-3) can be easily rearrange into the standard form

$$\dot{x}(t) = \boldsymbol{A}x(t) + \textstyle\sum_{i=1}^{N} \boldsymbol{A}_{di}x(t-\tau_i) \tag{8-4}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \bar{\boldsymbol{A}}_{11} & \cdots & \bar{\boldsymbol{A}}_{1N} \\ \vdots & \ddots & \vdots \\ \bar{\boldsymbol{A}}_{N1} & \cdots & \bar{\boldsymbol{A}}_{NN} \end{bmatrix}, \boldsymbol{A}_{di} = \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & & & \vdots \\ \mathbf{0} & & \bar{\boldsymbol{A}}_{dii} & & \mathbf{0} \\ \vdots & & & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix}.$$

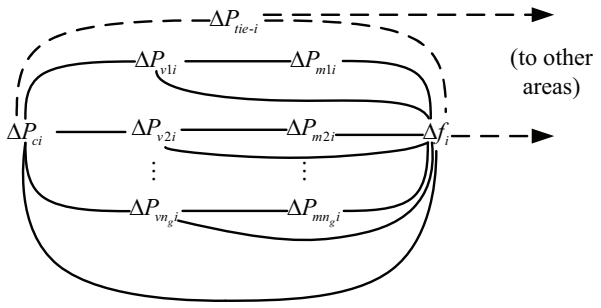


Figure 8-2 Connections of state variables of single-area (without dotted lines) and multi-area (with dotted lines) LFC.

8.2.2 Delay-dependent Stability Criterion

The major objectives of DDSA are to 1) calculate the system delay margins; 2) guide the controller design using the delay margins as an additional performance index. One of the mainstream methods of DDSA is based on Lyapunov-Krasovskii functionals (LKF) [59] whose existence implies the stability of time-delay systems. By specifying the structure of LKFs, sufficient conditions for stability of time-delay systems can be written as LMIs. This chapter will explain the structure-exploiting techniques based on the stability criterion derived from the results in [195].

Consider the linear system with a time-varying delay:

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{A}_d x(t-d(t)),\ t > 0 \\ x(t) = \phi(t),\ t \in [-\tau, 0] \end{cases} \tag{8-5}$$

where $x(t) \in \mathbb{R}^n$ is the state vector; $\boldsymbol{A}$ and $\boldsymbol{A}_d$ are system matrices with appropriate di-

mensions; the time-varying delay $d(t)$ is a continuous function satisfying $0 \leq d(t) \leq \tau$ and $|\dot{d}(t)| \leq \mu$; the initial condition $\phi(t)$ is a continuously differentiable function on $[-\tau, 0]$. By selecting the candidate LKF as

$$\begin{aligned} v(t) = x^T(t)\boldsymbol{P}x(t) + \int_{t-d(t)}^{t} x^T(s)\boldsymbol{Q}_1 x(s)ds + \\ \int_{t-\tau}^{t} x^T(s)\boldsymbol{Q}_2 x(s)ds + \tau \int_{-\tau}^{0}\int_{t+\theta}^{t} \dot{x}^T(s)\boldsymbol{R}\dot{x}(s)dsd\theta, \end{aligned} \tag{8-6}$$

the following theorem certifies the stability of system (8-5):

Theorem 8.1： System (8-5) is asymptotically stable if there exist symmetric positive-definite matrices $\boldsymbol{P} \succ \mathbf{0}$, $\boldsymbol{Q}_1 \succ \mathbf{0}$, $\boldsymbol{Q}_2 \succ \mathbf{0}$, $\boldsymbol{R} \succ \mathbf{0}$ and a appropriately dimensioned matrix $\boldsymbol{S}$ such that

$$\boldsymbol{\Psi} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{S} \\ \boldsymbol{S}^\top & \boldsymbol{R} \end{bmatrix} \succcurlyeq \mathbf{0} \tag{8-7}$$

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} & \boldsymbol{S} & \tau\boldsymbol{A}^\top\boldsymbol{R} \\ \boldsymbol{\Phi}_{12}^\top & \boldsymbol{\Phi}_{22} & -\boldsymbol{S}+\boldsymbol{R} & \tau\boldsymbol{A}_d^\top\boldsymbol{R} \\ \boldsymbol{S}^\top & -\boldsymbol{S}^\top+\boldsymbol{R} & -\boldsymbol{R}-\boldsymbol{Q}_2 & \mathbf{0} \\ \tau\boldsymbol{R}\boldsymbol{A} & \tau\boldsymbol{R}\boldsymbol{A}_d & \mathbf{0} & -\boldsymbol{R} \end{bmatrix} \prec 0 \tag{8-8}$$

where $\boldsymbol{\Phi}_{11} = \boldsymbol{P}\boldsymbol{A} + \boldsymbol{A}^\top\boldsymbol{P} + \boldsymbol{Q}_1 + \boldsymbol{Q}_2 - \boldsymbol{R}$, $\boldsymbol{\Phi}_{12} = \boldsymbol{P}\boldsymbol{A}_d + \boldsymbol{R} - \boldsymbol{S}$ and $\boldsymbol{\Phi}_{22} = -(1-\mu)\boldsymbol{Q}_1 - 2\boldsymbol{R} + \boldsymbol{S}^\top + \boldsymbol{S}$. □

Since the state-of-art LMI solvers have very limited ability to solve large-scale LMIs, Theorem 8.1 poses a great computational challenge to the application of DDSA to real-world systems. In the following sections, we describe the structure-exploiting techniques based on Theorem 8.1, but the proposed techniques are not restricted to Theorem 8.1 and can also be applied to other LMI based delay-dependent stability criteria.

8.3 Chordal-structured Matrix Decomposition

This section provides theory foundation for subsequent discussions. Some concepts and properties related to the chordal graph are reviewed at first. Then the theorem of chordal-structured matrix decomposition is described. Finally, several useful lemmas are presented and proved. The proofs can be skipped for pure application purpose.

8.3.1 Theoretical Background

Let $\mathbb{M}_n$ denote the set of $n \times n$ matrices, and the symmetric subset of $\mathbb{M}_n$ is $\mathbb{S}_n$. Every $\boldsymbol{A} \in \mathbb{M}_n$ is naturally associated with an undirected graph $G(V, E)$ where the vertex set $V =$

$\{1, 2, \cdots, n\}$ and the edge set $E = \{(i, j) \in V \times V : i \neq j, |[\boldsymbol{A}]_{ij}| + |[\boldsymbol{A}]_{ji}| \neq 0\}$ where $[\boldsymbol{A}]_{ij}$ denotes the $(i, j)^{\text{th}}$ element of $\boldsymbol{A}$. If $(i, j) \in E$, vertex i and j are said to be adjacent. The set of all vertices adjacent to vertex i is denoted by $adj_G(i)$. A graph is called complete if every pair of its vertices are adjacent. For any vertex subset $V' \subseteq V$, the subgraph induced by V' is a graph $G'(V', E')$ with $E' = E \cap (V' \times V')$. A clique is a subset of vertices of an undirected graph such that its induced subgraph is complete, and a clique is maximal if it is not a proper subset of another clique. Let $[n_1, n_2, \cdots, n_k]$ be a path of length k from vertex n_1 to vertex n_k, i.e. $(n_i, n_{i+1}) \in E$ for $1 \leq i \leq k-1$. Specially, a cycle is a path with $n_1 = n_k$. A graph is connected if there are paths containing each pair of vertices. A tree is an undirected connected graph with no cycles. In addition, a chord of a cycle is any edge joining two nonconsecutive vertices of the cycle.

The concept of chordal graph is central in this chapter:

Definition 8.1 ([174]): An undirected graph is chordal if every cycle of length greater than three has a chord. ◊

Other than the definition given above, the chordal graph can be characterized in several different ways [174]. The characterization of clique trees is of concern in this chapter. Let $G(V, E)$ be any graph. The set of all maximal cliques of G is denoted by $\mathcal{C} = \{C_1, C_2, \cdots, C_p\}$. Consider a tree $\mathcal{T}(\mathcal{C}, \mathcal{E})$ with vertices from $\mathcal{C}$ and edges from $\mathcal{E} \subseteq \mathcal{C} \times \mathcal{C}$. We call $\mathcal{T}(\mathcal{C}, \mathcal{E})$ a clique tree if it satisfies the clique-intersection property (**CIP**), i.e. for each pair of distinct maximal cliques $C_i, C_j \in \mathcal{C}$, the set $C_i \cap C_j$ is contained in every maximal clique on the path connecting C_i and C_j in the tree. In fact, the existence of a clique graph is equivalent to chordality:

Lemma 8.1 ([174]): A connected undirected graph G is chordal if and only if there exists a clique tree $\mathcal{T}(\mathcal{C}, \mathcal{E})$. □

On the other hand, given a graph $G(V, E)$, the set of matrices associated with this graph is denoted by $\mathbb{M}_n(E) = \{\boldsymbol{A} \in \mathbb{M}_n : [\boldsymbol{A}]_{ij} = 0 \text{ if } i \neq j \ \& \ (i, j) \notin E\}$. The symmetric subset of $\mathbb{M}_n(E)$ is denoted by $\mathbb{S}_n(E) = \mathbb{M}_n(E) \cap \mathbb{S}_n$. Further denote $\mathbb{S}_n^C = \{\boldsymbol{X} \in \mathbb{S}_n : [\boldsymbol{X}]_{ij} = 0 \text{ if } (i, j) \notin C \times C\}$ for every $C \subseteq V$. Let $\boldsymbol{E}_{ij}$ denote the appropriately dimensioned matrix with the $(i, j)^{\text{th}}$ element 1 and others 0.

Now we are ready to state the theorem of chordal-structured matrix decomposition as follows:

Theorem 8.2 ([196]): Given a chordal graph $G(V, F)$ with its maximal cliques $\mathcal{C} = \{C_1, C_2, \cdots, C_p\}$. $\mathcal{T}(\mathcal{C}, \mathcal{E})$ is a clique tree. Define $J(C) = \{(i, j) \in C \times C : i \leq j\}$ for every $C \subseteq V$ and

$$\Lambda = \{(g, h, k, l) : (g, h) \in J(C_k \cap C_l), (C_k, C_l) \in \mathcal{E}\}. \tag{8-9}$$

Any $\boldsymbol{A} \in \mathbb{S}_n(F)$ can be decomposed into $\tilde{\boldsymbol{A}}^k \in \mathbb{S}_n^{C_k}(k = 1, 2, \cdots, p)$ such that $\boldsymbol{A} = \sum_{k=1}^{p} \tilde{\boldsymbol{A}}^k$. Then $\boldsymbol{A} \succcurlyeq \boldsymbol{0}$ if and only if the system of LMIs

$$\tilde{\boldsymbol{A}}^k - \tilde{\boldsymbol{L}}^k(\boldsymbol{z}) \succcurlyeq \boldsymbol{0} \ (k = 1, 2, \cdots, p) \tag{8-10}$$

is feasible. Here $\boldsymbol{z} = (z_{ghkl} : (g, h, k, l) \in \Lambda)$ denotes a vetor variable consisting of $z_{ghkl}((g, h, k, l) \in \Lambda)$, and

$$\tilde{\boldsymbol{L}}^k(\boldsymbol{z}) = -\sum_{(i,j,h):(i,j,h,k)\in\Lambda} \boldsymbol{E}_{ij} z_{ijhk} + \sum_{(i,j,l):(i,j,k,l)\in\Lambda} \boldsymbol{E}_{ij} z_{ijkl} \tag{8-11}$$

for every $\boldsymbol{z} = (z_{ghkl} : (g, h, k, l) \in \Lambda)$. □

Note that the PSD condition on $\tilde{\boldsymbol{A}}^k - \tilde{\boldsymbol{L}}^k(\boldsymbol{z})$ is equivalent to the PSD condition on the submatrix indexed by C_k. Therefore, Theorem 8.2 transforms the PSD condition on a single $n \times n$ matrix into the PSD condition on multiple smaller matrices. When the sizes of the maximal cliques are much smaller than the size of $\boldsymbol{A}$, this transformation will bring significant computational advantages. For an arbitrary symmetric matrix $\boldsymbol{A}$, the corresponding graph $G(V, E)$ is not necessarily chordal. In this case, we can always find a chordal extension of $G(V, E)$, i.e. a chordal graph $G(V, F)$ with $F \supseteq E$, by adding some edges. Clearly, $\boldsymbol{A} \in \mathbb{S}_n(F)$. Then Theorem 8.2 can be readily applied to $G(V, F)$ and $\boldsymbol{A}$.

8.3.2 Useful Lemmas

Next, we present several lemmas which play vital roles in application of Theorem 8.2 to DDSA.

Lemma 8.2： Given an undirected chordal graph $G(V, F)$ with its maximum cliques $\mathcal{C} = \{C_1, C_2, \cdots, C_p\}$. For any symmetric matrix $\boldsymbol{A} \in \mathbb{S}_n(F)$ and matrix $\boldsymbol{B} \in \mathbb{M}_n(H)$ where

$$\begin{aligned} H = \{&(i, j) \in V \times V : i \neq j, \\ &\exists\, r,\ i, j \in C_r,\ adj_G(i) \setminus C_r = adj_G(j) \setminus C_r\}, \end{aligned} \tag{8-12}$$

$\boldsymbol{AB} \in \mathbb{M}_n(F)$ and $\boldsymbol{BA} \in \mathbb{M}_n(F)$. □

Proof： Matrix $\boldsymbol{B}$ can be written as $\boldsymbol{B} = \sum_{(i,j)\in H} b_{ij} \boldsymbol{E}_{ij}$ where $b_{ij} \in \mathbb{R}$. Therefore it suffices to show $\boldsymbol{AE}_{ij} \in \mathbb{M}_n(F)$ and $\boldsymbol{E}_{ij}\boldsymbol{A} \in \mathbb{M}_n(F)$, $\forall (i, j) \in H$. $\boldsymbol{AE}_{ij}$ is obtained by putting the i^{th} column of $\boldsymbol{A}$ in the j^{th} column of a $n \times n$ zero matrix. Let G' denote the graph associated with $\boldsymbol{AE}_{ij}$. From a graph perspective, we have

$$\begin{aligned} adj_{G'}(j) &= adj_G(i) \cup \{i\} \setminus \{j\} \\ &= (adj_G(i) \cap C_r) \cup (adj_G(i) \setminus C_r) \cup \{i\} \setminus \{j\} \\ &= (adj_G(j) \cap C_r) \cup (adj_G(j) \setminus C_r) \\ &= adj_G(j) \end{aligned} \tag{8-13}$$

where the first equality comes from the formation of $\boldsymbol{AE}_{ij}$ just described and the third equality comes from the completeness of clique C_r and the defining property of $(i,j) \in H$ in (8-12). From (8-13), we conclude $\boldsymbol{AE}_{ij} \in \mathbb{M}_n(F)$. Finally, $\boldsymbol{E}_{ij}\boldsymbol{A} \in \mathbb{M}_n(F)$ results from $\boldsymbol{A}^T\boldsymbol{E}_{ji} \in \mathbb{M}_n(F)$ noticing the undirectness of graph G. ■

Remark 8.1： Lemma 8.2 characterizes the structure of $\boldsymbol{B}$ allowing $\boldsymbol{AB}$ and $\boldsymbol{BA}$ to inherit the chordal sparsity pattern of $\boldsymbol{A}$. In fact, two categories of elements (i,j) satisfy the defining property stated in (8-12). First, i and j are inner vertices of the same maximal clique C_r, i.e. $adj_G(i) \subseteq C_r$ and $adj_G(j) \subseteq C_r$. Second, vertex i and j are adjacent to the same vertices outside the clique they both belong to. If nonzero elements of $\boldsymbol{B}$ are only allowed at (i,j) from above two categories, the multiplication $\boldsymbol{AB}$ will not create new nonzero elements outside F thus inherit the sparsity of $\boldsymbol{A}$. □

Lemma 8.3： Given m chordal graphs $G^{(k)}(V^{(k)}, F^{(k)})$, $k = 1, 2, \cdots, m$ of exactly the same structure, the maximum cliques of the k^{th} graph are denoted by $\mathcal{C}^{(k)} = \{C_1^{(k)}, C_2^{(k)}, \cdots, C_p^{(k)}\}$. Another graph $G(V,F)$ is constructed as follows:

$$V = \cup_{k=1}^{m} V^{(k)} \tag{8-14}$$

$$F = \{(i,j) \in \cup_{r=1}^{p} (C_r \times C_r) : i \neq j\} \tag{8-15}$$

where $C_r = \cup_{k=1}^{m} C_r^{(k)}$. Then $G(V,F)$ is a chordal graph with $\mathcal{C} = \{C_1, C_2, \cdots, C_p\}$ being the maximum cliques. □

Proof： Since $G^{(k)}(V^{(k)}, F^{(k)})$ is chordal, there exists a clique tree $\mathcal{T}^{(k)}(\mathcal{C}^{(k)}, \mathcal{E}^{(k)})$, $\forall k = 1, 2, \cdots, m$ according to Lemma 8.1, and all the clique trees can be chosen with exactly the same structure. Let $\mathcal{T}(\mathcal{C}, \mathcal{E})$ denote the tree on $\mathcal{C}$ analogous to $\mathcal{T}^{(k)}(\mathcal{C}^{(k)}, \mathcal{E}^{(k)})$. According to Lemma 8.1, it suffices to show $\mathcal{T}(\mathcal{C}, \mathcal{E})$ is a clique tree of graph $G(V,F)$. We first show $\mathcal{C} = \{C_1, C_2, \cdots, C_p\}$ are maximum cliques. It is explicitly shown in (8-15) that C_r, $\forall r = 1, \cdots, p$, is a clique of $G(V,F)$. To prove the maximality, assume $C_r' \supset C_r$ is another clique. Pick one vertex $v \in C_r' \setminus C_r$. We have $v \in V^{(k)}$ for some $k \in \{1, 2, \cdots, m\}$. Then $C_r \cap V^{(k)} \supset C_r^{(k)}$ is also a clique of $G^{(k)}(V^{(k)}, F^{(k)})$, which contradicts with the maximality of $C_r^{(k)}$. We then show the CIP holds for $\mathcal{T}(\mathcal{C}, \mathcal{E})$. For any $C_i, C_j \in \mathcal{C}$,

$$\begin{aligned} C_i \cap C_j &= \left(\cup_{k=1}^{m} C_i^{(k)}\right) \cap \left(\cup_{k=1}^{m} C_j^{(k)}\right) \\ &= \cup_{k=1}^{m} \left(C_i^{(k)} \cap C_j^{(k)}\right). \end{aligned} \tag{8-16}$$

For any C_r on the path connecting C_i and C_j, $C_r^{(k)}$ is also on the path connecting $C_r^{(i)}$ and $C_r^{(j)}$ due to the similarity of $\mathcal{T}(\mathcal{C}, \mathcal{E})$ and $\mathcal{T}^{(k)}(\mathcal{C}^{(k)}, \mathcal{E}^{(k)})$. Since $\mathcal{T}^{(k)}(\mathcal{C}^{(k)}, \mathcal{E}^{(k)})$ is a clique tree, $C_i^{(k)} \cap C_j^{(k)} \subseteq C_r^{(k)}$. Using (8-16), we have $C_i \cap C_j \subseteq \cup_{k=1}^{m} C_r^{(k)} = C_r$. Therefore, CIP holds for $\mathcal{T}(\mathcal{C}, \mathcal{E})$. We finally conclude $\mathcal{T}(\mathcal{C}, \mathcal{E})$ is a clique tree of $G(V, F)$. ■

Remark 8.2： Lemma 8.3 states that a large chordal graph can be constructed from a group of small chordal graphs with the same structure by merging corresponding maximum cliques of small chordal graphs. □

Lemma 8.3 is useful when dealing with the block matrix with each block having the same chordal sparsity pattern. The following corollary is a direct consequence of Lemma 8.3.

Corollary 8.1： Let

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_{11} & \cdots & \mathbf{\Phi}_{1m} \\ \vdots & \ddots & \vdots \\ \mathbf{\Phi}_{m1} & \cdots & \mathbf{\Phi}_{mm} \end{bmatrix} \tag{8-17}$$

be a $nm \times nm$ symmetric matrix with each block $\mathbf{\Phi}_{ij} \in \mathbb{M}_n$. $G(V, F)$ is a chordal graph with vertices $V = \{1, \cdots, n\}$ and maximum cliques $\mathcal{C} = \{C_1, C_2, \cdots, C_p\}$. If $\mathbf{\Phi}_{ij} \in \mathbb{M}_n(F), \forall 1 \leq i, j \leq m$, then $\mathbf{\Phi}$ is associated with chordal graph $G'(V', F')$, i.e. $\mathbf{\Phi} \in \mathbb{S}_{nm}(F')$ where $V' = \cup_{k=0}^{m-1}(kn + V)$ and $F' = \{(i, j) \in \cup_{r=1}^{p}(\cup_{k=0}^{m-1}(kn + C_r)) \times (\cup_{k=0}^{m-1}(kn + C_r)) : i \neq j\}$. $G'(V', F')$ is a chordal graph with

$$C_r' = \cup_{k=0}^{m-1}(kn + C_r), r = 1, \cdots, p \tag{8-18}$$

being the maximal cliques. □

Based on Corollary 1, we then discuss a more complex situation with the $\mathbf{\Phi}_{11}$ block in (8-17) being fully dense and other blocks following the same chordal sparsity pattern.

Lemma 8.4： Let $\mathbf{\Phi}$ be the $nm \times nm$ block symmetric matrix shown in (8-17) where $\mathbf{\Phi}_{11}$ is fully dense. $\mathbf{\Phi}_{ij},\ i + j > 2$, is associated with graph $G(V, F)$, i.e. $\mathbf{\Phi}_{ij} \in \mathbb{M}_n(F), \forall i + j > 2$. $G(V, F)$ is a chordal graph with vertices $V = \{1, \cdots, n\}$ and maximal cliques $\mathcal{C} = \{C_1, C_2, \cdots, C_p\}$. Define $\hat{V} = \cup_{1 \leq r, t \leq p}(C_r \cap C_t)$. Then $\mathbf{\Phi}$ is associated with chordal graph $G^*(V^*, F^*)$, i.e. $\mathbf{\Phi} \in \mathbb{S}_{nm}(F^*)$ where $V^* = \cup_{k=0}^{m-1}(kn + V)$ and $F^* = \{(i, j) \in \cup_{r=0}^{p} C_r^* \times C_r^* : i \neq j\}$ with

$$C_0^* = V \cup \left(\cup_{k=1}^{m-1}(kn + \hat{V})\right) \tag{8-19}$$

$$C_r^* = \cup_{k=0}^{m-1}(kn + C_r),\ r = 1, \cdots, p. \tag{8-20}$$

Accordingly, $\mathcal{C}^* = \{C_0^*, C_1^*, \cdots, C_p^*\}$ are the maximal cliques of $G^*(V^*, F^*)$. □

Proof： Step 1: we show $\mathbf{\Phi} \in \mathbb{S}_{nm}(F^*)$, i.e. $(i,j) \in F^*$ if $[\mathbf{\Phi}]_{ij} \neq 0$. It suffices to consider three cases. First, note that the submatrix of $\mathbf{\Phi}$

$$\begin{bmatrix} \mathbf{\Phi}_{22} & \cdots & \mathbf{\Phi}_{2m} \\ \vdots & \ddots & \vdots \\ \mathbf{\Phi}_{m2} & \cdots & \mathbf{\Phi}_{mm} \end{bmatrix}$$

satisfies the conditions of Corollary 1, which implies that $\forall n+1 \leq i,j \leq nm$ with $[\mathbf{\Phi}]_{ij} \neq 0$, $\exists 1 \leq r \leq p$, such that $(i,j) \in (\cup_{k=1}^{m-1}(kn + C_r)) \times (\cup_{k=1}^{m-1}(kn + C_r))$, leading to $(i,j) \in C_r^* \times C_r^*$. Second, $\forall 1 \leq i \leq n < j \leq nm$ with $[\mathbf{\Phi}]_{ij} \neq 0$, we have $(i, j \bmod n) \in F$ according to the chordal structure of each block other than $\mathbf{\Phi}_{11}$. Hence $\exists 1 \leq r \leq p$, such that $(i, j \bmod n) \in C_r \times C_r$, which implies $(i,j) \in C_r^* \times C_r^*$ by definition. Third, $\forall 1 \leq i,j \leq n$, it is obvious that $(i,j) \in C_0^* \times C_0^*$.

Step 2: we prove $G^*(V^*, F^*)$ is chordal by showing the maximality of cliques $\mathcal{C}^* = \{C_0^*, C_1^*, \cdots, C_p^*\}$ and the existence of a clique tree $\mathcal{T}^*(\mathcal{C}^*, \mathcal{E}^*)$. First, the maximality of cliques $\mathcal{C}^*$ relies on the observation that $C_r^* \nsubseteq C_t^*$, $\forall\ 0 \leq r,t \leq p$ by definition. Second, we define $\mathcal{T}^*(\mathcal{C}^*, \mathcal{E}^*)$ a tree on $\mathcal{C}^*$ as depicted in Fig. 8-3. To check CIP holds for $\mathcal{T}^*(\mathcal{C}^*, \mathcal{E}^*)$, it suffices

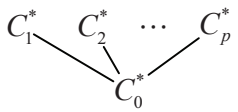


Figure 8-3 Clique Tree of Chordal Graph $G^*(V^*, F^*)$

to show $C_r^* \cap C_t^* \subseteq C_0^*$, $\forall\ 1 \leq r < t \leq p$. Note that

$$\begin{aligned} C_r^* \cap C_t^* &= \left(\cup_{k=0}^{m-1}(kn + C_r)\right) \cap \left(\cup_{k=0}^{m-1}(kn + C_t)\right) \\ &= \cup_{k=0}^{m-1}\left(kn + C_r \cap C_t\right) \\ &\subseteq \cup_{k=0}^{m-1}\left(kn + \hat{V}\right) \\ &\subseteq C_0^*. \end{aligned} \tag{8-21}$$

The two inclusions in (8-21) come from the definitions of $\hat{V}$ and C_0^*, respectively. Therefore, $\mathcal{T}^*(\mathcal{C}^*, \mathcal{E}^*)$ is a clique tree. Using Lemma 8.1, we conclude $G^*(V^*, F^*)$ is chordal.

Step 1 and step 2 together lead to the lemma. ■

Remark 8.3： The structure of $\mathbf{\Phi}$ described in Lemma 8.4 is what will be encountered in the DDSA. If the sizes of $\hat{V}$ and $C_r, r = 1, \cdots, p$ are much smaller than n, the sizes of $C_r^*, r = 0, \cdots, p$ would be much smaller than nm. Then Theorem 8.2 can be employed to decompose the PSD condition on $\mathbf{\Phi}$ into PSD conditions on much smaller matrices. □

8.4 Structure-exploiting Delay-dependent Stability Analysis of Load Frequency Control

This section presents the techniques to exploit chordal sparsity and symmetry in DDSA of LFC.

8.4.1 Exploiting Chordal Sparsity

In practical problems, like power system LFC, the system matrices $\boldsymbol{A}$ and $\boldsymbol{A}_d$ exhibit strong sparsity. We introduce the aggregate system matrix

$$\bar{\boldsymbol{A}} = |\boldsymbol{A}| + \mathbf{1}_n|\boldsymbol{A}_d| \tag{8-22}$$

where $\mathbf{1}_n$ denotes the $n \times n$ all-ones matrix, and the aggregate sparsity pattern graph $G(V, E)$ with $V = \{1, 2, \cdots, n\}$ and $E = \{(i, j) \in V \times V : i \neq j, |[\bar{\boldsymbol{A}}]_{ij}| + |[\bar{\boldsymbol{A}}]_{ji}| \neq 0\}$. The sparsity of the system matrices is thus represented by the sparsity of graph $G(V, E)$. However, the sparsity of the system matrices do not readily lead to the sparsity of the LMI (8-7) and (8-8) due to the full flexibility of weighting matrices. To take advantages of the sparsity of system matrices, we need to properly restrict the structure of weighting matrices $\boldsymbol{Q}_1$, $\boldsymbol{Q}_2$, $\boldsymbol{R}$ and $\boldsymbol{S}$ so that the LMI (8-7) and (8-8) can inherit the sparsity from system matrices. This idea is made precise by the following theorem.

Theorem 8.3： Let $G(V, F)$ be a chordal extension of the aggregate sparsity pattern graph $G(V, E)$ of system (8-5) as defined above. $\mathcal{C} = \{C_1, C_2, \cdots, C_p\}$ are the maximal cliques of $G(V, F)$. H is the edge set defined in (8-12). Let $\boldsymbol{P} \in \mathbb{S}_n$, $\boldsymbol{Q}_1, \boldsymbol{Q}_2 \in \mathbb{S}_n(F)$, $\boldsymbol{R} \in \mathbb{S}_n(H)$ and $\boldsymbol{S} \in \mathbb{M}_n(F)$. Then $\boldsymbol{\Psi}$ (8-7) is associated with the chordal graph $G'(V', F')$ defined in Corollary 1 with $m = 2$, and $\boldsymbol{\Phi}$ (8-8) is associated with the chordal graph $G^*(V^*, F^*)$ defined in Lemma 8.4 with $m = 4$. □

Proof： According to Lemma 8.2, $\boldsymbol{Q}_1, \boldsymbol{Q}_2 \in \mathbb{S}_n(F)$, $\boldsymbol{R} \in \mathbb{S}_n(H)$ and $\boldsymbol{S} \in \mathbb{M}_n(F)$ imply that all blocks of $\boldsymbol{\Phi}$ except $\boldsymbol{\Phi}_{11}$ are in $\mathbb{M}_n(F)$, and $\boldsymbol{\Phi}_{11}$ is fully dense in general, i.e. $\boldsymbol{\Phi}$ satisfies the conditions in Lemma 8.4. Therefore, $\boldsymbol{\Phi} \in \mathbb{S}_{4n}(F^*)$ where the chordal graph $G^*(V^*, F^*)$ is defined in Lemma 4 with $m = 4$. Moreover, since each block of $\boldsymbol{\Psi}$ belongs to $\mathbb{M}_n(F)$, $\boldsymbol{\Psi} \in \mathbb{S}_{2n}(F')$ with the chordal graph $G'(V', F')$ defined with $m = 2$ in Corollary 8.1. ■

Remark 8.4： As the consequence of Theorem 8.3, all matrices except $\boldsymbol{P}$ in Theorem 8.1 with PSD conditions are associated with related chordal graphs, i.e. $\boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{R} \in \mathbb{S}_n(F)$, $\boldsymbol{\Psi} \in \mathbb{S}_{2n}(F')$ and $\boldsymbol{\Phi} \in \mathbb{S}_{4n}(F^*)$. Therefore, Theorem 8.2 can be employed to decompose these PSD conditions into smaller ones. □

The effects of enforcing chordal sparsity on weighting matrices as described in Theorem

8.3 are twofold. First, the number of decision variables is reduced; Second, the sizes of the PSD conditions are reduced. Both effects contribute to the improvement of numerical tractability. However, by restricting the structure of weighting matrices $\boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{R}$ and $\boldsymbol{S}$, the obtained stability criterion becomes more conservative than Theorem 8.1, which is the price has to be paid for improving the numerical tractability. The conservatism brought by the structure restriction of weighting matrices will be assessed by numerical studies in section 8.5. Note that we do not enforce chordal sparsity on $\boldsymbol{P}$ due to the experience that enforcing chordal sparsity on $\boldsymbol{P}$ will make the stability criterion overly conservative thus useless in practice. This is in stack contrast to the pure control theory paper [194] which enforces sparsity on $\boldsymbol{P}$ and exploits the sparsity of $\boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}$.

Consider the single-area LFC with turbine governor model shown in Fig. 8-1 as an example. It can be verified by definition that the graph related to $\bar{\boldsymbol{A}}$ of single-area LFC shown in Fig. 8-4 is already a chordal graph without adding edges, i.e. $F = E$. The maximal cliques of this chordal graph are $\mathcal{C}_i = \{\Delta P_c, \Delta f, \Delta P_{vi}, \Delta P_{mi}\}, i = 1, 2, \cdots, n$. The edges that belong to set H defined in (8-12) are shown with red dotted lines. By restricting the structures of weighting matrices as stated in Theorem 8.3, the $(2n_g + 2) \times (2n_g + 2)$ PSD conditions on $\boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{R}$ can all be reduced to 4×4 PSD conditions on their submatrices indexed by the maximal cliques. The PSD conditions on Ψ and Φ can also be reduced according to the maximal cliques of $G'(V', F')$ and $G^*(V^*, F^*)$ given in Corollary 1 and Lemma 8.4, respectively.

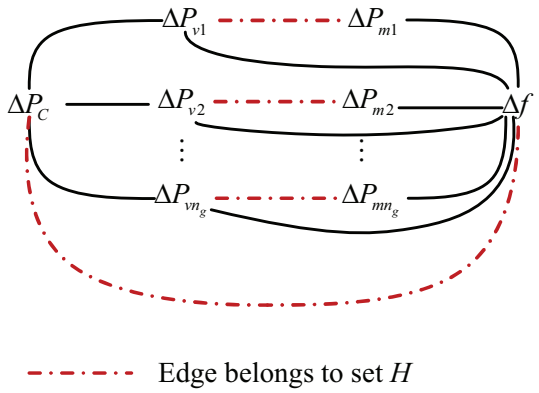


Figure 8-4 Chordal graph for single-area LFC. The maximal cliques are $\mathcal{C}_i = \{\Delta P_c, \Delta f, \Delta P_{vi}, \Delta P_{mi}\}, i = 1, 2, \cdots, n$.

8.4.2 Exploiting Symmetry

Due to the full flexibility of weighting matrix $\boldsymbol{P}$, the DDSA is still computationally intensive even when chordal sparsity is exploited. We need to further restrict the structure of $\boldsymbol{P}$ by exploiting symmetry.

The basic idea rests on the observation that many elements of matrix $\boldsymbol{P}$ take very similar

values when we directly conduct DDSA using Theorem 8.1 on a small-scale LFC system. The fundamental reason for this phenomenon is the highly symmetric structure of LFC control loops shown in Fig. 8-2 and similar parameters of generation units. Note that $\boldsymbol{P}$ is the weighting matrix related to the first term in LKF (8-6). If the "nature" of the state variable pair (i, j) is the same as that of state variable pair (k, l), $[\boldsymbol{P}]_{ij}$ and $[\boldsymbol{P}]_{kl}$ should take the similar values. For example, in a single-area LFC, assume that unit i and unit j have similar parameters. In the LKF of the system, the coefficients of $\Delta f \Delta P_{vi}$ and $\Delta f \Delta P_{vj}$ should take similar values because ΔP_{vi} and ΔP_{vj} are of the same relation to Δf. It is also the case for $\Delta f \Delta P_{mi}$ and $\Delta f \Delta P_{mj}$, $\Delta P_c \Delta P_{vi}$ and $\Delta P_c \Delta P_{vj}$, $\Delta P_c \Delta P_{mi}$ and $\Delta P_c \Delta P_{mj}$, etc. Moreover, we conjecture that although the parameters of generation units are different in practice, enforcing the above-mentioned symmetry on $\boldsymbol{P}$ would not introduce much conservatism to the results of DDSA.

To formalize the idea described above, we need to introduce the notion of permutation group. A permutation σ is a bijective map from set $\{1, 2, \cdots, n\}$ to itself. For our propose, here n is the dimension of the system matrix. A set of permutations Σ is called a permutation group if it is closed under composition and contains the identity map. The permutation group of concern to us is constructed as follows:

1. Endow the set of all generation units $\{g_1, \cdots, g_{n_g}\}$ with an equivalence relation $\sim$. We say $g_h \sim g_k$ if and only if they belong to the same control area and their turbine-governor systems have similar structure and parameters.
2. Let $\{i^1_{g_k}, \cdots, i^p_{g_k}\}$ be the indices of the state variables related to the turbine-governor system of generation unit g_k, and $\mathcal{I}_g = \bigcup_{\forall k,l}\{i^l_{g_k}\}$. Define the permutation group Σ as the set of all permutations satisfying $\sigma(i) = i$ if $i \notin \mathcal{I}_g$ and $\forall g_k$, $\exists g_h \sim g_k$ such that $\sigma(i^l_{g_k}) = i^l_{g_h}, \forall 1 \leq l \leq p$.

The definition of the permutation group given above allows us to establish a equivalence relation on pairs of state variables: $(i, j) \sim (k, l)$ if and only if there exists $\sigma \in \Sigma$ such that $k = \sigma(i)$ and $l = \sigma(j)$. Hence the vague idea of "state variable pair (i, j) and (k, l) are of the same nature" is precisely characterized by the equivalence relation $(i, j) \sim (k, l)$. As a result, $[\boldsymbol{P}]_{ij}$ and $[\boldsymbol{P}]_{\sigma(i)\sigma(j)}$ are expected to take the same value, i.e. $[\boldsymbol{P}]_{ij} = [\boldsymbol{P}]_{\sigma(i)\sigma(j)}, \forall \sigma \in \Sigma$ and we say $\boldsymbol{P}$ is Σ-invariant. To precisely characterize the structure of $\boldsymbol{P}$, we need to introduce the concept of the orbit. The orbit to which the pair (i, j) belong under the permutation group Σ is given by

$$\mathcal{O}(i, j) = \{(\sigma(i), \sigma(j)) | \sigma \in \Sigma\}. \tag{8-23}$$

It is well known that the orbits partition the set on which the group operates [197], i.e. the set $\{1, \cdots, n\} \times \{1, \cdots, n\}$ is partitioned into several orbits $\mathcal{O}_1, \mathcal{O}_2, \cdots, \mathcal{O}_r$. Based on the definition of the permutation group Σ, we explicitly show the structures of the orbits in Table I. Matrix $\boldsymbol{P}$ thus takes the same value on each orbit. For every $k \in \{1, \cdots, r\}$ we define

Table 8-1 Orbits of Group Σ Operating on Index Pair (i,j)

(i,j)	$\mathcal{O}(i,j)$
$i,j \notin \mathcal{I}_g$	$\{(i,j)\}$
$(i, i^l_{g_k}),\ i \notin \mathcal{I}_g$	$\{(i, i^l_{g_h}) : g_h \sim g_k\}$
$(i^l_{g_k}, i),\ i \notin \mathcal{I}_g$	$\{(i^l_{g_h}, i) : g_h \sim g_k\}$
$(i^l_{g_k}, i^m_{g_k})$	$\{(i^l_{g_h}, i^m_{g_h}) : g_h \sim g_k\}$
$(i^l_{g_k}, i^m_{g_h}), g_k \neq g_h$	$\{(i^l_{g_{k'}}, i^m_{g_{h'}}) : g_{k'} \sim g_k, \sim g_{h'} \sim g_h, k' \neq h'\}$

$\hat{\boldsymbol{B}}_k \in \{0,1\}^{n\times n}$ by $[\hat{\boldsymbol{B}}_k]_{ij} = 1$ if $(i,j) \in \mathcal{O}_k$ and $[\hat{\boldsymbol{B}}_k]_{ij} = 0$ otherwise. Then $\hat{\boldsymbol{B}}_1, \cdots, \hat{\boldsymbol{B}}_r$ form a basis of the space of Σ-invariant matrices (not necessarily symmetric). The basis of the space of symmetric Σ-invariant matrices can then be obtained by setting $\boldsymbol{B}_k = \hat{\boldsymbol{B}}_h + \hat{\boldsymbol{B}}_l$ if $\hat{\boldsymbol{B}}_h^T = \hat{\boldsymbol{B}}_l$ and $\boldsymbol{B}_k = \hat{\boldsymbol{B}}_k$ if $\hat{\boldsymbol{B}}_k^T = \hat{\boldsymbol{B}}_k$. Therefore, $\boldsymbol{P}$ is parametrized by $\boldsymbol{P} = \sum_{k=1}^{p} p_k \boldsymbol{B}_k$.

To exemplify the effects of the above procedure, we first consider a single-area LFC with three generation units. The turbine governor models of all generating units are as shown in Fig. 8-1 and we assume all generating units are all similar to each other. Then parametrized matrix $\boldsymbol{P} = \sum_{k=1}^{13} p_k \boldsymbol{B}_k$ is depicted in Fig. 8-5. The number of decision variables are reduced from $(2n_g+2)(2n_g+3)/2$ to 13.

$$
\begin{array}{c}
\\
\Delta P_c \\
\Delta f \\
\Delta P_{v1} \\
\Delta P_{v2} \\
\Delta P_{v3} \\
\Delta P_{m1} \\
\Delta P_{m2} \\
\Delta P_{m3}
\end{array}
\begin{array}{c}
\begin{array}{cccccccc}
\Delta P_c & \Delta f & \Delta P_{v1} & \Delta P_{v2} & \Delta P_{v3} & \Delta P_{m1} & \Delta P_{m2} & \Delta P_{m3}
\end{array} \\
\begin{bmatrix}
p_1 & p_2 & p_3 & p_3 & p_3 & p_4 & p_4 & p_4 \\
p_2 & p_5 & p_6 & p_6 & p_6 & p_7 & p_7 & p_7 \\
p_3 & p_6 & p_8 & p_9 & p_9 & p_{10} & p_{11} & p_{11} \\
p_3 & p_6 & p_9 & p_8 & p_9 & p_{11} & p_{10} & p_{11} \\
p_3 & p_6 & p_9 & p_9 & p_8 & p_{11} & p_{11} & p_{10} \\
p_4 & p_7 & p_{10} & p_{11} & p_{11} & p_{12} & p_{13} & p_{13} \\
p_4 & p_7 & p_{11} & p_{10} & p_{11} & p_{13} & p_{12} & p_{13} \\
p_4 & p_7 & p_{11} & p_{11} & p_{10} & p_{13} & p_{13} & p_{12}
\end{bmatrix}
\end{array}
$$

Figure 8-5 Structure of the symmetric Σ-invariant matrix $\boldsymbol{P}$ for a single-area LFC with three generation units

The same procedure can be readily applied to three-area LFC. We consider a three-area LFC scheme with three generation units in each area and generation units in each area are assumed to be similar to each other. To simplify the presentation, we partition the weighting matrix $\boldsymbol{P}$ into six independent blocks, i.e.

$$
\boldsymbol{P} = \begin{bmatrix}
\boldsymbol{P}_{11} & \boldsymbol{P}_{12} & \boldsymbol{P}_{13} \\
\boldsymbol{P}_{12}^T & \boldsymbol{P}_{22} & \boldsymbol{P}_{23} \\
\boldsymbol{P}_{13}^T & \boldsymbol{P}_{23}^T & \boldsymbol{P}_{33}
\end{bmatrix} \tag{8-24}
$$

where $\boldsymbol{P}_{ii} (i = 1,2,3)$ represents the coefficients of quadratic term within the control area i,

and $\boldsymbol{P}_{ij}$ denotes the coefficients of the cross-product term between area i and area j. Since the structure of $\boldsymbol{P}_{ii}$ is similar to that of single-area LFC, we only show the structure of $\boldsymbol{P}_{ij}$ in Fig. 8-6. The number of decision variables in $\boldsymbol{P}_{ij}$ is reduced from $(2n_g+3)^2$ to 25.

$$
\begin{array}{c}
\\
\Delta P_{ci} \\
\Delta f_i \\
\Delta P_{\text{tie-}i} \\
\Delta P_{v1i} \\
\Delta P_{v2i} \\
\Delta P_{v3i} \\
\Delta P_{m1i} \\
\Delta P_{m2i} \\
\Delta P_{m3i}
\end{array}
\begin{array}{c}
\begin{array}{ccccccccc}
\Delta P_{cj} & \Delta f_j & \Delta P_{\text{tie-}j} & \Delta P_{v1j} & \Delta P_{v2j} & \Delta P_{v3j} & \Delta P_{m1j} & \Delta P_{m2j} & \Delta P_{m3j}
\end{array} \\
\begin{bmatrix}
p_1 & p_2 & p_3 & p_4 & p_4 & p_4 & p_5 & p_5 & p_5 \\
p_6 & p_7 & p_8 & p_9 & p_9 & p_9 & p_{10} & p_{10} & p_{10} \\
p_{11} & p_{12} & p_{13} & p_{14} & p_{14} & p_{14} & p_{15} & p_{15} & p_{15} \\
p_{16} & p_{17} & p_{18} & p_{19} & p_{19} & p_{19} & p_{20} & p_{20} & p_{20} \\
p_{16} & p_{17} & p_{18} & p_{19} & p_{19} & p_{19} & p_{20} & p_{20} & p_{20} \\
p_{16} & p_{17} & p_{18} & p_{19} & p_{19} & p_{19} & p_{20} & p_{20} & p_{20} \\
p_{21} & p_{22} & p_{23} & p_{24} & p_{24} & p_{24} & p_{25} & p_{25} & p_{25} \\
p_{21} & p_{22} & p_{23} & p_{24} & p_{24} & p_{24} & p_{25} & p_{25} & p_{25} \\
p_{21} & p_{22} & p_{23} & p_{24} & p_{24} & p_{24} & p_{25} & p_{25} & p_{25}
\end{bmatrix}
\end{array}
$$

Figure 8-6 Structure of the symmetric Σ-invariant partial matrix $\boldsymbol{P}_{ij}$ for a three-area LFC with three generation units in each area

The symmetry-exploiting technique described above significantly reduces the number of decision variables, which contributes to the improvement of numerical tractability. However, the structure restriction of weighting matrix $\boldsymbol{P}$ will also bring additional conservatism which will be assessed by numerical studies in section 8.5.

8.4.3 Implementation Issues

The main steps of **structure-exploiting DDSA** are summarized as follows:

Step1. Obtain a chordal extension $G(V,F)$ and its maximal cliques of the aggregate sparsity pattern graph. This can be accomplished by performing a Cholesky factorization to $\bar{\boldsymbol{A}}+\bar{\boldsymbol{A}}^T+\epsilon\boldsymbol{I}$ (Here ϵ is any positive number to guarantee the positive definiteness of the matrix to be factorized) with approximate minimum degree ordering, i.e. $\bar{\boldsymbol{A}}+\bar{\boldsymbol{A}}^T+\epsilon\boldsymbol{I}=\boldsymbol{L}^T\boldsymbol{L}$. The sparsity pattern of the Cholesky factor defines the chordal extension and the maximal cliques, i.e. $F=\{(i,j)\in V\times V: i\neq j, |[\boldsymbol{L}]_{ij}|+|[\boldsymbol{L}]_{ji}|\neq 0\}$.

Step2. Obtain the edge set H defined in (8-12) and restrict the structure of $\boldsymbol{Q}_1$, $\boldsymbol{Q}_2$, $\boldsymbol{R}$ and $\boldsymbol{S}$ as: $\boldsymbol{R},\boldsymbol{Q}_1,\boldsymbol{Q}_2\in\mathbb{S}_n(H)$ and $\boldsymbol{S}\in\mathbb{M}_n(H)$. Note that although Theorem 8.3 only requires $\boldsymbol{Q}_1,\boldsymbol{Q}_2\in\mathbb{S}_n(F)$ and $\boldsymbol{S}\in\mathbb{M}_n(F)$, we found that further restricting $\boldsymbol{Q}_1,\boldsymbol{Q}_2\in\mathbb{S}_n(H)$ and $\boldsymbol{S}\in\mathbb{M}_n(H)$ does not bring much conservatism in practice. Under above structure restriction of the weighting matrices, Theorem 8.3 guarantees the existence of chordal sparsity in the LMIs of Theorem 8.1.

Step3. Identify the equivalence relation among generation units and the permutation group Σ on the indices of state variables, and find out the orbits $\mathcal{O}_1,\mathcal{O}_2,\cdots,\mathcal{O}_r$ according to Table 8-1.

Step4. Find out the basis $\boldsymbol{B}_1, \cdots, \boldsymbol{B}_r$ of all possible weighting matrix $\boldsymbol{P}$ with desirable symmetry structure according to the orbits of permutation group Σ. Then $\boldsymbol{P}$ is restricted as $\boldsymbol{P} = \sum_{k=1}^{p} p_k \boldsymbol{B}_k$.

Step5. Form the LMIs in Theorem 8.1 and reformulate it with the chordal-structured matrix decomposition (Theorem 8.2). This step can be implemented conveniently using existing software package SparseCoLO which automatically detects the chordal structure in general semi-definite programming (SDP) and reformulates the SDP to facilitate the solution using standard solvers [196].

Step6. Solve the obtained new LMIs with standard LMI solvers.

It it worth mentioning that the above procedure is totally algorithmic and does not reply on specific models of turbine-governor systems and AGC controllers. Hence the proposed method is adaptive to different models employed. Fig. 8-7 shows the difference between the proposed structure-exploiting DDSA and the direct DDSA. In the structure-exploiting DDSA, the stability criterion is pre-processed by the sparsity and symmetry exploiting techniques before solved by the standard LMI solver. Whereas the stability criterion is directly sent to the solver without any processing in the direct DDSA. The practical effects of the pre-processing will be comprehensively assessed by numerical studies in the next section.

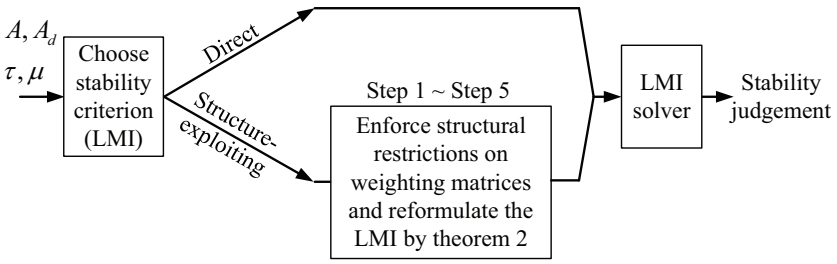


Figure 8-7 Procedures of direct and structure-exploiting DDSA

8.5 Numerical Results

This section presents the numerical studies of the proposed structure-exploiting DDSA with comparison to the direct DDSA. The computational performance and introduced conservatism of the proposed structure-exploiting techniques are analyzed in details. The methods are implemented in MATLAB 2015b with YALMIP [118] as the modeling tool and SDPT3 [198] as the solver. SparseCoLO [196] is used to automates the chordal-structured matrix decomposition. The program runs on a Win8 PC with a 3.0 GHz CPU and 24 GB RAM.

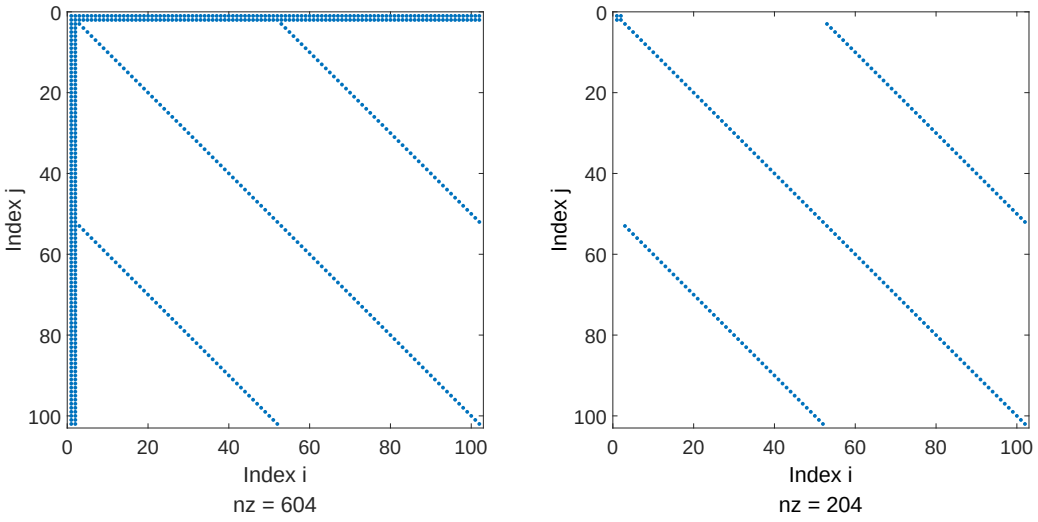
The numerical tests are based on 10 unit 39 bus New England system (NE39). Larger test systems are constructed by directly scale up the NE39. For example, the 200 unit system is obtained by merging 20 NE39 systems. The generator inertia data can be found in [199] and the generator rated power and load demand are extracted from MATPOWER [116]. The

load damping D in each synchronous area is assumed to be 1% total load/Hz. The typical values for droop characteristic R of each governor vary from 3% to 7% p.u./rated power. The turbine-governor system of each unit is modelled as in Fig. 8-1. The typical parameters for turbine-governor systems are $T_g = 0.08s$ and $T_t = 0.40s$ [61]. In every numerical test, the actual values for T_g, T_t and R are randomly generated in the range of $1\pm\chi\%$ of the typical values given above. The value of χ varies from 0 to 50 in the tests to reflect the different degree of non-symmetry from parameter variation. The participation factors are also randomly selected between 0 and 1. Three-area test cases are obtained by connecting three single-area systems. The connection parameters are $T_{12} = 0.20$, $T_{13} = 0.25$, $T_{21} = 0.20$, $T_{23} = 0.12$, $T_{32} = 0.25$ and $T_{32} = 0.12$.

8.5.1 Computational Efficiency

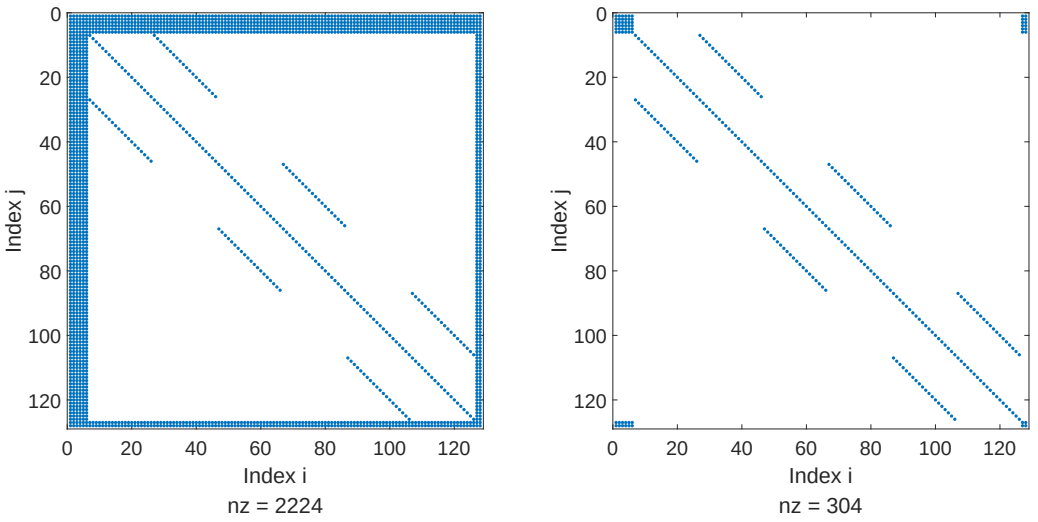
The proposed structure-exploiting DDSA has been employed in both single-area and three-area LFC schemes. For single-area problems, the structure-exploiting techniques are applied to Theorem 8.1. For three-area problems, Theorem 8.1 is no longer applicable since three independent time delays exist in the whole system and Theorem 8.1 can only tackle single time delay. In this case, we apply the structure-exploiting technique to the stability criterion dealing with multiple constant delays proposed in our previous paper [58]. Fig. 8-8 shows the sparsity pattern of matrices in $\mathbb{S}_n(F)$ and $\mathbb{S}_n(H)$ decribed in Theorem 8.3 for single-area LFC with 50 generation units, and Fig. 8-9 depicts the sparsity pattern of matrices in $\mathbb{S}_n(F)$ and $\mathbb{S}_n(H)$ for three-area LFC with 20 generation units in each area. Both figures give us intuitive ideas about the sparsity of LFC loops and the significant reduction in decision variables in weighting matrices when sparsity is exploited.

In the first group of tests, we fix the typical value of R to be 5% and set $\chi = 25$. Structure-exploiting DDSA and direct DDSA are conducted on LFC systems with different numbers of generation units. The computational performance statistics are summarized in Table 8-2. Column 2 and 3 denote the numbers of generation units and the orders of the systems, respectively. Four statistics are used to assess the scale of the LMIs to be solved with both methods. **no. psd** and **max psd** denote the number and maximal size of PSD constraints, respectively. Since LMIs are transformed into standard SDP forms before solving by conic program solvers, the numbers of equality constraints (**m**) and decision variables (**n**) in standard primal form [198] are also taken as indicators of the problem scales. Moreover, **solver time** in Table 8-2 denotes the CPU time in seconds for SDPT3 to solve corresponding problems. As shown in Table 8-2, by exploiting the chordal sparsity in DDSA, the numbers of PSD constraints increase whereas the sizes of PSD constraints decrease. Note that Theorem 8.3 only proves the existence rather than the uniqueness of the chordal structure in the LMIs after restricting the structure of the weighting matrices. SparseCoLO automatically selects a "good" chordal extension by merging



(a) $\mathbb{S}_n(F)$ (b) $\mathbb{S}_n(H)$

Figure 8-8 Sparsity pattern of matrices in $\mathbb{S}_n(F)$ (left) and $\mathbb{S}_n(H)$ (right) described in Theorem 8.3 for single-area LFC with 50 generation units



(a) $\mathbb{S}_n(F)$ (b) $\mathbb{S}_n(H)$

Figure 8-9 Sparsity pattern of matrices in $\mathbb{S}_n(F)$ (left) and $\mathbb{S}_n(H)$ (right) described in Theorem 8.3 for three-area LFC with 20 generation units in each area

some maximal cliques in the cliques tree to balance the increase in the number and decrease in the size of PSD constraints [200]. In general, by exploiting both chordal sparsity and symmetry, the decrease of m and n showed in Table 8-2 indicates the significant reduction in the scales of LMIs, which is further reflected in the solver time of each problem. For both single-area and multi-area LFC, direct DDSA is numerically tractable for systems with less than around 40

Table 8-2 Comparison of Computational Performance of the Direct and Structure-exploiting DDSA

	ng	ord.	Direct DDSA					Stucture-exploiting DDSA				
			no. psd	max psd	m	n	solver time (s)	no. psd	max psd	m	n	solver time (s)
1-area	30	62	6	248	11656	92256	387.77	157	66	876	12734	2.03
	50	102	6	408	31416	249696	inf. (out of mem.)	257	106	1436	29154	4.51
	150	202	6	808	122816	979296	inf. (out of mem.)	507	206	2836	98204	19.03
	200	402	6	1608	485616	3878496	inf. (out of mem.)	1007	406	5636	356304	89.11
	500	1002	6	4008	3014016	24096096	inf. (out of mem.)	2507	1002	14036	2090604	827.32
3-area	3*10	68	8	476	16422	258944	810.77	280	100	1738	22917	5.88
	3*20	128	8	896	57792	917504	inf. (out of mem.)	551	146	3225	54449	30.30
	3*40	248	8	1736	216132	3444224	inf. (out of mem.)	1091	266	5954	164262	136.61
	3*60	368	8	2576	475272	7583744	inf. (out of mem.)	1631	386	8640	331484	404.87

generation units, whereas structure-exploiting DDSA applies to systems with hundreds of generation units. For problems solvable by both methods, the solver time of structure-exploiting DDSA is two orders of magnitude less than that of direct DDSA. In a real-world power system, there could be hundreds of generation units. The improvement of the numerical tractability of the DDSA method would allow more detailed modeling for each control area, which increases the accuracy and reliability of the results of DDSA.

8.5.2 Evaluation of Additional Conservatism

To exploit chordal sparsity and symmetry in DDSA, structure restrictions are enforced on the weighting matrices, which inevitably brings additional conservatism. If too much conservatism is introduced, the method could become useless in practice. In the second group of tests, we evaluate the introduced conservatism by comparing the delay margins (stable delay region) obtained by the structure-exploiting DDSA and the direct DDSA on relatively small-scale systems. The typical value of R and χ are also set to be 5% and 25, respectively. Table. 8-3 reports the comparison of delay margins by two methods on single-area LFC with ten generation units under different parameters of the PI controller. The results show that the delay margins obtained by the structure-exploiting DDSA ($\boldsymbol{\tau_{se}}$) are uniformly larger than 90% of the delay margins by the direct DDSA ($\boldsymbol{\tau_{dr}}$), which indicates that the structure-exploiting techniques only introduce minor conservatism. Table 8-4 lists the side lengths of the cubic stable delay region of three-area LFC obtained by both methods. For presentation simplicity, each control area is assumed to have the same PI parameters. The results in Table 8-4 also confirm the conclusion that the structure-exploiting techniques only introduce minor conservatism. These tests numerically validate the conjecture we raised in section 8.4.2. To better understand the underlying mechanism, we compare the mesh plots of obtained weighting matrices $\boldsymbol{P}$ and $\boldsymbol{Q}_1$ by two methods in Fig. 8-10 and Fig. 8-11. Both figures show that the overall structure of weighting matrices obtained by the structure-exploiting DDSA looks very similar to that of weighting matrices obtained by the direct DDSA. In other words, the proposed structure restrictions on weighting matrices are not very restrictive, so the structure-exploiting techniques only bring minor conservatism. Fig. 8-12 shows the comparison of frequency deviation for

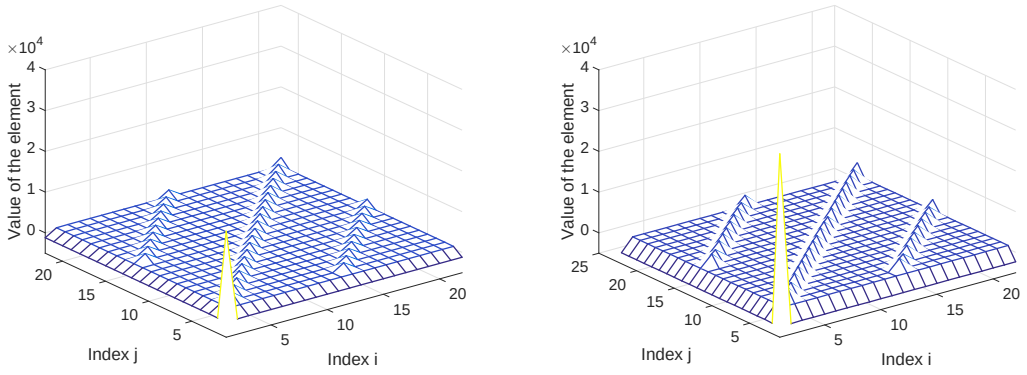
Table 8-3 Comparison of Delay Margins (seconds) by Two Methods on Single-area LFC with 10 Generation Units

(K_P, K_I)	τ_{se}	τ_{dr}	ratio
(0.05,0.05)	27.1765	28.5269	95.27%
(0.05,0.1)	13.4079	13.8895	96.53%
(0.05,0.15)	8.8376	8.9960	98.24%
(0.1,0.05)	26.3379	27.8265	94.65%
(0.1,0.1)	12.9913	14.1550	91.78%
(0.1,0.15)	8.5794	9.2047	93.21%

Table 8-4 Comparison of length of the Cubic Stable Delay Region (seconds) by Two Methods on Three-area LFC with 10 Generation Units in Each Area

(K_P, K_I)	τ_{se}	τ_{dr}	ratio
(0.05,0.05)	25.4993	26.4615	96.36%
(0.05,0.1)	12.5912	12.7927	98.43%
(0.05,0.15)	8.2242	8.2516	99.67%
(0.1,0.05)	24.6634	25.7419	95.81%
(0.1,0.1)	12.1820	13.0518	93.34%
(0.1,0.15)	8.0786	8.4567	95.53%

0.1 p.u. step load disturbance of single-area LFC with τ_{se} and τ_{dr} in Table 8-3. As shown in the time-domain simulation, both τ_{se} and τ_{dr} are conservative approximation to the underlying true delay margins, but τ_{se} possesses a little more conservatism than τ_{dr}.

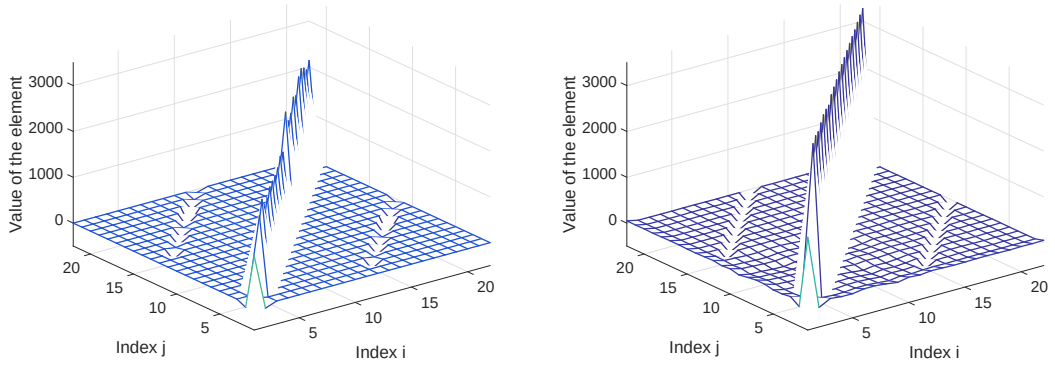


(a) Structure-exploiting DDSA (b) Direct DDSA

Figure 8-10 Comparison of mesh plots of $\boldsymbol{P}$ obtained by two methods on single-area LFC with 10 generation units

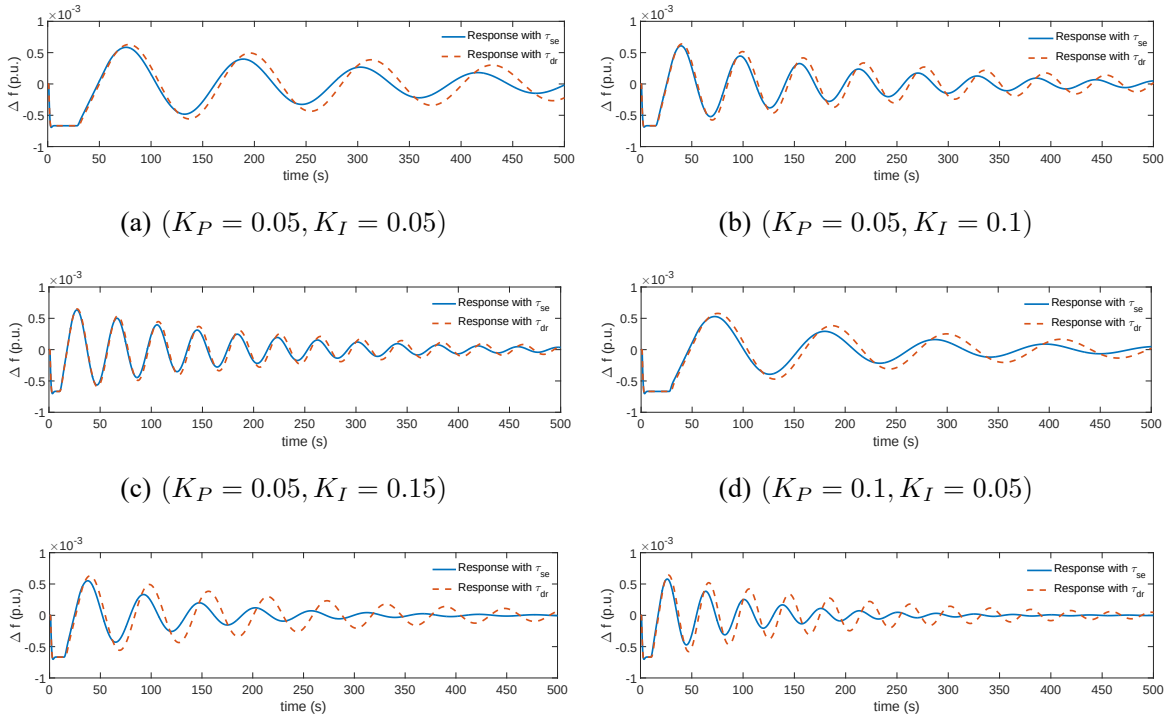
In the following, we further define the degree of additional conservatism as

$$\rho_c = 1 - \frac{\tau_{se}}{\tau_{dr}}. \tag{8-25}$$



(a) Structure-exploiting DDSA (b) Direct DDSA

Figure 8-11 Comparison of mesh plots of $\boldsymbol{Q}_1$ obtained by two methods on single-area LFC with 10 generation units



(a) $(K_P = 0.05, K_I = 0.05)$ (b) $(K_P = 0.05, K_I = 0.1)$

(c) $(K_P = 0.05, K_I = 0.15)$ (d) $(K_P = 0.1, K_I = 0.05)$

(e) $(K_P = 0.1, K_I = 0.1)$ (f) $(K_P = 0.1, K_I = 0.15)$

Figure 8-12 Comparison of system Frequency response of the 10-unit single-area LFC with the delay margins obtained by two methods

In the third group of tests, we investigate the sensitivity of ρ_c to the parameter $\chi\%$ and the typical value of R. The tests are performed on 10-unit single-area LFC. The value of $\chi\%$ reflects the degree of the non-symmetry of the control loops. The larger $\chi\%$ is, the more non-symmetric the system is and the more conservatism could be introduced by symmetry-exploiting technique. Fig. 8-13 shows the relationship between ρ_c and $\chi\%$ which confirms our qualitative analysis. It is shown ρ_c increases quite slow as $\chi\%$ increases and ρ_c is less

than 12% even when system parameters T_g, T_t and R are subject to maximal deviation of 50% of related typical values. In other words, the symmetry-exploiting is quite robust to the non-symmetry originated from parameter deviation. From our numerical experience, we found ρ_c was quite sensitive to the typical value of droop characteristics R. Fig. 8-14 reveals the quantitative relationship between ρ_c and typical value of R. It is observed that the smaller the typical value of R is, the more conservatism will be introduced by the structure-exploiting techniques. In fact, the smaller R is, the stronger the connection between Δf and ΔP_{vk} is. The sparsity-exploiting technique brings more conservatism when the connection between Δf and ΔP_{vk} becomes stronger. In general, compared with its significant speed-up and scale-up effects, the proposed structure-exploiting techniques introduce reasonably acceptable degree of conservatism.

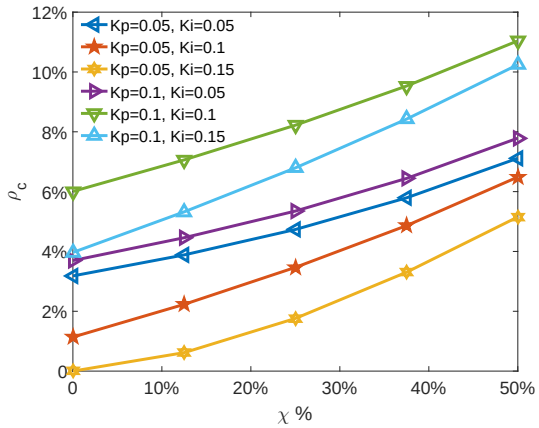


Figure 8-13 Degree of additional conservatism of structure-exploiting techniques versus $\chi\%$. Typical value of R is set to be 5%.

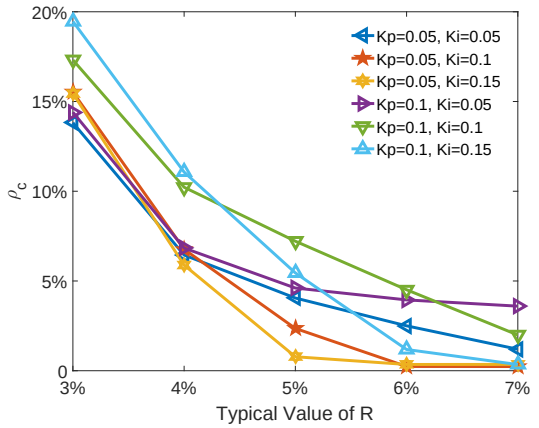


Figure 8-14 Degree of additional conservatism of structure-exploiting techniques versus the typical Value of R. $\chi\%$ is set to be 25%.

8.6 Discussion

Although the main focus of this chapter is stability analysis, the proposed method can also be extended to controller design. The controller should not only guarantee system stability but also provide desirable dynamic performance. In the framework of robust control, the dynamic performance is represented by some robust performance index (RPI), e.g. H_∞ norm $\sup_{\omega\neq 0}\frac{\|z\|_2}{\|\omega\|_2}$ where ω is the disturbance and z is the controlled output. In LFC, ω usually represents load disturbance and z often includes frequency deviation and area control errors [61]. Every delay-dependent stability criterion can be extended to the corresponding delay-dependent BRL which provides guarantee for both system stability and dynamic performance thus can be used to design an appropriate controller. For example, the BRL corresponding to Theorem 8.1 is stated as follows: consider a system expressed by

$$\begin{cases} \dot{x}(t) = \boldsymbol{A}x(t) + \boldsymbol{A}_d x(t-d(t)) + \boldsymbol{B}_\omega \omega(t) \\ z(t) = \boldsymbol{C}_\omega x(t) \end{cases} \tag{8-26}$$

where matrix $\boldsymbol{A}_d$ is a function of the design parameter (K_P, K_I). Then the system (8-26) is asymptotically stable and $\sup_{\omega\neq 0}\frac{\|z\|_2}{\|\omega\|_2} < \gamma$ for any delay $0 \leq d(t) \leq \tau$ and $|\dot{d}(t)| \leq \mu$ if there exist $\boldsymbol{P} \succ \boldsymbol{0}$, $\boldsymbol{Q}_1 \succ \boldsymbol{0}$, $\boldsymbol{Q}_2 \succ \boldsymbol{0}$, $\boldsymbol{R} \succ \boldsymbol{0}$ and $\boldsymbol{S}$ such that $\boldsymbol{\Psi} \succcurlyeq \boldsymbol{0}$ and

$$\begin{bmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} & \boldsymbol{S} & \tau\boldsymbol{A}^\top\boldsymbol{R} & \boldsymbol{P}\boldsymbol{B}_\omega & \boldsymbol{C}_\omega^T \\ \boldsymbol{\Phi}_{12}^\top & \boldsymbol{\Phi}_{22} & -\boldsymbol{S}+\boldsymbol{R} & \tau\boldsymbol{A}_d^\top\boldsymbol{R} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{S}^\top & -\boldsymbol{S}^\top+\boldsymbol{R} & -\boldsymbol{R}-\boldsymbol{Q}_2 & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \tau\boldsymbol{R}\boldsymbol{A} & \tau\boldsymbol{R}\boldsymbol{A}_d & \boldsymbol{0} & -\boldsymbol{R} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{B}_\omega^T\boldsymbol{P} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & -\gamma\boldsymbol{I}_n & \boldsymbol{0} \\ \boldsymbol{C}_\omega & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & -\gamma\boldsymbol{I}_n \end{bmatrix} \prec 0 \tag{8-27}$$

where $\boldsymbol{\Psi}$, $\boldsymbol{\Phi}_{11}$, $\boldsymbol{\Phi}_{12}$ and $\boldsymbol{\Phi}_{22}$ are defined as in Theorem 8.1. As shown above, the BRL is also a set of LMIs very similar to the stability criterion. Therefore, the sparsity and symmetry exploiting techniques proposed in this chapter can be readily applied to the BRL to accelerate and scale up the control design.

More specifically, based on the framework in [201], we can design a delay-dependent H_∞ robust controller which minimizes the RPI (H_∞ norm) while maintaining stability for any delays less than a preset upper bound. By simply replacing Lemma 8.1 in [201] with the above-mentioned dealy-dependent BRL, a numerical method to design such controller is readily available. Note that the LMI needs to be solved iteratively for probably hundreds of times to converge to the optimal controller parameters; hence the accelerating and scaling-up effects provided by the proposed structure-exploiting techniques could be even more significant.

In addition, LMI-based delay-dependent stability criteria are employed to calculate the delay margins of wide-area damping controllers (WADC) installed at generator excitation sys-

tems [**?**] and FACTS devices [56]. Delay-dependent BRLs are also employed in [**?**] to design state-feedback WADC. Therefore, the structure-exploiting techniques proposed in this chapter are also of potential to be applied to the analysis and synthesis of the time-delay WADC.

8.7 Conclusion

In this chapter, the chordal sparsity and symmetry of the graph related to time-delay LFC loops have been exploited to improve the numerical tractability of DDSA. The graph-theoretic analysis provides guidance for restricting the structure of weighting matrices in DDSA, such that the LMIs possess chordal sparsity. The symmetry of LFC loops has been utilized to reduce the number of decision variables. At the price of introducing minor conservatism, case studies show the numerical tractability and computational efficiency have been improved by orders of magnitude. Lemma 8.2 in this chapter is a generalization of the results in [194]. Our result provides a general approach for $\boldsymbol{A}^T\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}$ to inherit the chordal sparsity of $\boldsymbol{A}$, whereas paper [194] discusses several special cases. Note that the conservatism introduced by structure-exploiting techniques are dependent on system parameters. One future direction is to derive theoretical bounds for the introduced conservatism for a certain class of systems. Another future direction is to extend the proposed method to DDSA of WADC.

9 Path-following Method to Design Power System Damping Controllers Considering Time Delays

9.1 Introduction

Power systems are interconnected electromechanical systems whose normal operation depends on the synchronization of the connected generators. The syncronized state is always challenged by the continuous load/generation changes and sporadic faults. To resist those disturbances, the power systems must have enough damping. Power system stabilizers (PSS) are designed and widely implemented to improve the damping of power systems [111]. Conventional PSSs using local feedback signals are very effective to damp local oscillation modes. As the systems become more and more interconnected, low-frequency inter-area oscillation becomes prominent issues and wide-area damping control (WADC) using remote feedback signals are then developed to provide enough damping for inter-area modes [202, 203]. The remote signals are sent through communication networks and are unavoidably subject to time delays. Therefore, time delays need to be considered in the design of damping control systems involving WADCs.

Two groups of methods have been applied to analyze the stability and performance of time-delay WADC. The first group is the frequency-domain method based on eigenanalysis [44–47]. To obtain the eigenvalues, the time-delay system is often approximated by an augmented delay-free linear system through discretization of the continuous state function of the time-delay system. Usually, the dimension of the delay-free system is much higher than the time-delay system and proportional to the number of discretization points employed. As the number of discretization points N increases, the eigenvalues of the delay-free system converge to those of the time-delay system at the rate of $\mathcal{O}(N^{-N})$. Hence, practically speaking, frequency-domain methods can obtain the exact eigenvalues along with other accurate performance metrics, e.g. delays margins, damping ratios and damping factors, for systems with constant delays. The classical frequency-domain phase compensation approaches [48–50] have been the most practically applicable methods to design PSSs and FACTS controllers to suppress oscillations. However, their extensions to time-delay power systems are scarcely reported.

The second group of methods is the time-domain methods [51–55] which leverage linear matrix inequalities (LMI) as the computational tools. In time-domain methods, the stability of the time-delay systems is certified by the existence of certain Lyapunov-Krasovskii functionals (LKF). To find such LKFs, a candidate LKF with undetermined weighting matrices is first constructed, and some integral inequalities are employed to bound the time derivative of the LKF. Then the positive definiteness of the LKF and the negative definiteness of its deriva-

tive are then expressed as LMIs. The time-domain methods are always conservative, but the conservatism can be reduced by using more general LKF and tighter integral inequalities. In addition, they can deal with both constant and time-varying delays and facilitate the robust design of the controllers. When it comes to control synthesis, it is easy for time-domain methods to formulate the problems due to its affinity to modern H_∞ and H_2 control theory. However, it is not always easy to actually solve those problems. Only when the full state feedback control is adopted, the problem can be equivalently formulated as convex semi-definite programs (SDP) thus solvable by off-the-shelf solvers [53]. For other control architectures, the problem is bilinear [68] or more general nonlinear SDPs which are non-convex and NP-hard. Aside from the tractability issue, almost all works of LMI-based methods use reduced-order models of the power systems, due to the very limited ability of state-of-the-art SDP solvers to solve large-scale problems. Although from a computational perspective, it is desired to work with the reduced models, any assertion on the stability of the reduced system does not necessarily imply the same result for the original full-order system. In fact, the reduced system may prettify the stability of the full-order system, which makes the analysis and synthesis unreliable.

This chapter further develops time-domain methods for analysis and synthesis of damping control systems involving time delays. To remedy the deficiency of using the reduced model, we cast the full-order system into the feedback interconnection of a time-delay reduced close-loop system and a time-delay error system. Considering the H_∞ error of model reduction, we establish a condition to ensure the ϵ-exponential stability of the full-order system only using the reduced close-loop system model, by leveraging tools from dissipativity theory. Then we formulate the damping control design as a nonlinear SDP minimizing a carefully defined H_2 performance metric. A path-following method is proposed to solve the problem to local optimality. Tests on several benchmark systems show the proposed method is very effective to significantly improve system damping by the coordinated design of PSSs and WADCs.

9.2 System Model

The power system electromechanical dynamics are usually described by a set of differential algebraic equations [111]. After linearizing the equations at a given operation equilibrium and eliminating the algebraic variables, the model is cast into a linear dynamical system described by ordinary differential equations. The state-space representation of the linear system is given by:

$$\begin{cases} \dot{\boldsymbol{x}}_s = \boldsymbol{A}_s \boldsymbol{x}_s + \boldsymbol{B}_s \boldsymbol{u}_s \\ \boldsymbol{y}_s = \boldsymbol{C}_s \boldsymbol{x}_s \end{cases} \tag{9-1}$$

Linear system (9-1) captures the local dynamics of the power system at given operation point and are widely used to study small-signal stability and design damping controllers. To improve

system damping, we attempt to install w wide-area and l local controllers in the systems. The state-space representation of the ith controller is:

$$\begin{cases} \dot{\boldsymbol{x}}_{ci} = \boldsymbol{A}_{ci}\boldsymbol{x}_{ci} + \boldsymbol{b}_{ci}u_{ci} \\ y_{ci} = \boldsymbol{c}_{ci}^{\top}\boldsymbol{x}_{ci} + d_{ci}u_{ci} \end{cases} \tag{9-2}$$

The wide-area controllers employ remote signals as inputs and the signals need to be transmitted through the communication networks, therefore the feedback signals of wide-area controllers are inevitably time-delayed. In contrast, the local controllers use locally available signals and the communication delays are negligible. Fig. 9-1 sketches out the power system damping control system involving both local and wide-area controllers. Considering the com-

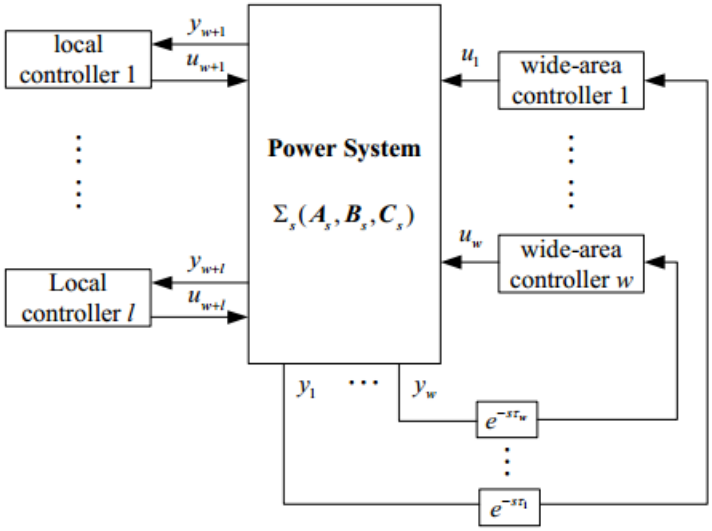


Figure 9-1 Power System with Local and Wide-area Damping Controllers

munication delays of wide-area signals, the close-loop system is then given by

$$\begin{bmatrix} \dot{\boldsymbol{x}}_s \\ \dot{\boldsymbol{x}}_c \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{A}}_{11} & \hat{\boldsymbol{A}}_{12} \\ \hat{\boldsymbol{A}}_{21} & \hat{\boldsymbol{A}}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_s \\ \boldsymbol{x}_c \end{bmatrix} + \sum_{i=1}^{w} \hat{\boldsymbol{A}}_{\tau i}\boldsymbol{x}_s(t-\tau_i) \tag{9-3}$$

where

$$\hat{\boldsymbol{A}}_{11} = \boldsymbol{A}_s + \sum_{i=w+1}^{w+l} \boldsymbol{b}_{si}d_{ci}\boldsymbol{c}_{si}^{\top} \tag{9-4}$$

$$\hat{\boldsymbol{A}}_{12} = \sum_{i=1}^{w+l} \boldsymbol{b}_{si}\boldsymbol{c}_{ci}^{\top}\boldsymbol{S}_{ci}, \quad \hat{\boldsymbol{A}}_{21} = \sum_{i=w+1}^{w+l} \boldsymbol{S}_{ci}^{\top}\boldsymbol{b}_{ci}\boldsymbol{c}_{si}^{\top} \tag{9-5}$$

$$\hat{\boldsymbol{A}}_{22} = \sum_{i=1}^{w+l} \boldsymbol{S}_{ci}^{\top}\boldsymbol{A}_{ci}\boldsymbol{S}_{ci}, \quad \hat{\boldsymbol{A}}_{\tau i} = \begin{bmatrix} \boldsymbol{b}_{si}d_{ci}\boldsymbol{c}_{si}^{\top} \\ \boldsymbol{S}_{ci}^{\top}\boldsymbol{b}_{ci}\boldsymbol{c}_{si}^{\top} \end{bmatrix} \tag{9-6}$$

and $\boldsymbol{S}_{ci}$ is a connection matrix such that $\boldsymbol{x}_{ci} = \boldsymbol{S}_{ci}\boldsymbol{x}_c$; $\boldsymbol{c}_{si}^{\top}$ is the ith row of $\boldsymbol{C}_s$ and $\boldsymbol{b}_{si}$ is the ith column of $\boldsymbol{B}_s$.

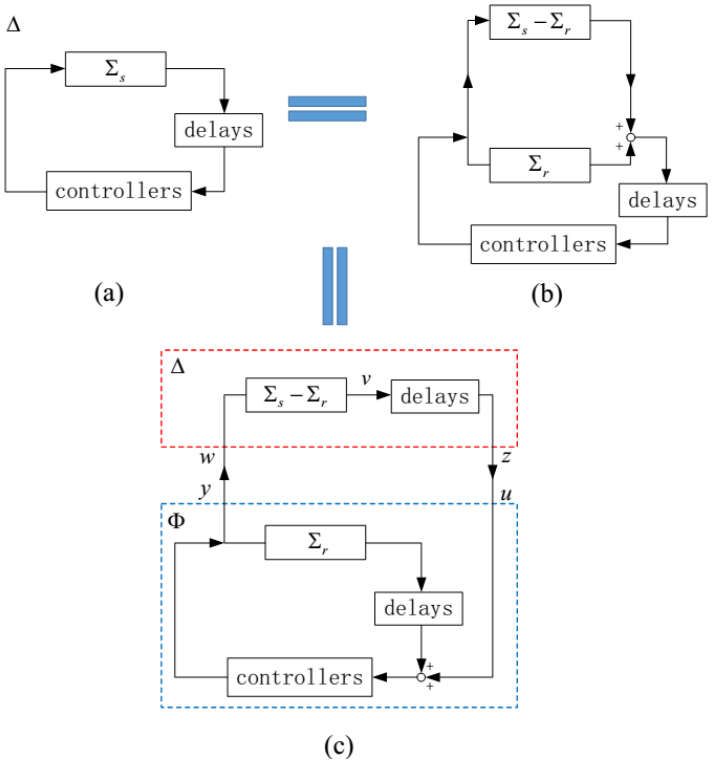


Figure 9-2 Illustration of Model Reduction

To alleviate computational burden, the designers often replace the full-order power system model Σ_s with a reduced-order model given by

$$\begin{cases} \dot{\boldsymbol{x}}_r = \boldsymbol{A}_r \boldsymbol{x}_r + \boldsymbol{B}_r \boldsymbol{u}_r \\ \boldsymbol{y}_r = \boldsymbol{C}_r \boldsymbol{x}_r \end{cases} \tag{9-7}$$

Then the orignal close-loop system is strictly equivalent to the feedback interconnection of the reduced close-loop system Φ and the delayed reduction error system Δ, illustrated in Fig. 9-2. The state-space representation of the reduced close-loop system Φ is then written as

$$\begin{bmatrix} \dot{\boldsymbol{x}}_r \\ \dot{\boldsymbol{x}}_c \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_r \\ \boldsymbol{x}_c \end{bmatrix} + \sum_{i=1}^{w} \boldsymbol{A}_{\tau i} \boldsymbol{x}_r (t - \tau_i) + \boldsymbol{B}\boldsymbol{u} \tag{9-8}$$

$$\boldsymbol{y} = \boldsymbol{C}^{\top} \begin{bmatrix} \boldsymbol{x}_r \\ \boldsymbol{x}_c \end{bmatrix} + \sum_{i=1}^{w} \boldsymbol{C}_{\tau i} \boldsymbol{x}_r (t - \tau_i) + \boldsymbol{D}\boldsymbol{u} \tag{9-9}$$

where $\boldsymbol{A}_{ij}, \boldsymbol{A}_{\tau i}$ $(1 \le i, j \le 2)$ share the same formula as in (9-4)∼(9-6) by replacing the subscript s with subscript r;

$$\boldsymbol{B} = \begin{bmatrix} \sum_{i=1}^{w+l} \boldsymbol{b}_{ri} d_{ci} \boldsymbol{e}_i^{\top} \\ \sum_{i=1}^{w+l} \boldsymbol{S}_{ci}^{\top} \boldsymbol{b}_{ci} \boldsymbol{e}_i^{\top} \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} \sum_{i=w+1}^{w+l} \boldsymbol{c}_{ri} d_{ci} \boldsymbol{e}_i^{\top} \\ \sum_{i=1}^{w+l} \boldsymbol{S}_{ci}^{\top} \boldsymbol{c}_{ci} \boldsymbol{e}_i^{\top} \end{bmatrix} \tag{9-10}$$

Table 9-1 Definition of Notations

$\boldsymbol{\Theta}(\epsilon)=\sum_{i=1}^{w}\boldsymbol{A}_{\eta i}^{\top}\boldsymbol{P}_i\boldsymbol{J}_i+\boldsymbol{J}_i^{\top}\boldsymbol{P}_i\boldsymbol{A}_{\eta i}+2\epsilon\boldsymbol{J}_i^{\top}\boldsymbol{P}_i\boldsymbol{J}_i+\mathrm{e}^{2\epsilon\tau_i}\boldsymbol{E}_{0r}^{\top}\boldsymbol{R}_{0i}\boldsymbol{E}_{0r}-\boldsymbol{F}_{1i}^{\top}\boldsymbol{R}_{0i}\boldsymbol{F}_{1i}+\tau_i\mathrm{e}^{2\epsilon\tau_i}\boldsymbol{E}_{0r}^{\top}\boldsymbol{R}_{1i}\boldsymbol{E}_{0r}-\boldsymbol{\Pi}_i+\frac{\tau_i^2}{2}\mathrm{e}^{2\epsilon\tau_i}\boldsymbol{E}_{0r}^{\top}\boldsymbol{R}_{2i}\boldsymbol{E}_{0r}-\boldsymbol{\Psi}_i$
$\boldsymbol{E}_0=[\boldsymbol{I}_{n+m},\boldsymbol{0}_{(n+m)\times 4nw}],\ \ \boldsymbol{E}_{0r}=[\boldsymbol{I}_n,\boldsymbol{0}_{n\times m}]\boldsymbol{E}_0,\ \ \boldsymbol{E}_i=[\boldsymbol{0}_{(4n)\times(n+m+4n(i-1))},\boldsymbol{I}_{4n},\boldsymbol{0}_{4n\times 4n(w-i)}],\ i=1,\cdots,w$
$\boldsymbol{J}_i=\mathrm{col}\{\boldsymbol{E}_0,\boldsymbol{F}_{2i},\boldsymbol{F}_{3i},\boldsymbol{F}_{4i}\},\ \ \boldsymbol{F}_j=[\boldsymbol{0}_{n\times(j-1)n},\boldsymbol{I}_n,\boldsymbol{0}_{n\times(4-j)n}],\ \boldsymbol{F}_{ji}=\boldsymbol{F}_j\boldsymbol{E}_i,\ j=1,\cdots,4,\ i=1,\cdots,w$
$\boldsymbol{A}_\tau=\boldsymbol{A}\boldsymbol{E}_0+\sum_{i=1}^{w}\boldsymbol{A}_{\tau i}\boldsymbol{F}_{1i},\ \ \boldsymbol{B}_\tau=\mathrm{col}\{\boldsymbol{B},\boldsymbol{0}_{3n\times(w+l)}\},\ \ \boldsymbol{C}_\tau^{\top}=\boldsymbol{C}^{\top}\boldsymbol{E}_0+\sum_{i=1}^{w}\boldsymbol{C}_{\tau i}^{\top}\boldsymbol{F}_{1i}$
$\boldsymbol{A}_{\eta i}=\mathrm{col}\{\boldsymbol{A}_\tau,\mathrm{e}^{\epsilon\tau_i}\boldsymbol{E}_{0r}-\boldsymbol{F}_{1i}-\epsilon\boldsymbol{F}_{2i},\tau_i\mathrm{e}^{\epsilon\tau_i}\boldsymbol{E}_{0r}-\boldsymbol{F}_{2i}-\epsilon\boldsymbol{F}_{3i},\frac{\tau_i^2}{2}\mathrm{e}^{\epsilon\tau_i}\boldsymbol{E}_{0r}-\boldsymbol{F}_{3i}-\epsilon\boldsymbol{F}_{4i}\},\ i=1,\cdots,w$
$\boldsymbol{\Pi}_i=\frac{1}{\tau_i}\boldsymbol{F}_{2i}^{\top}\boldsymbol{R}_{1i}\boldsymbol{F}_{2i}+\frac{3}{\tau_i}(\boldsymbol{F}_{2i}-\frac{2}{\tau_i}\boldsymbol{F}_{3i})^{\top}\boldsymbol{R}_{1i}(\boldsymbol{F}_{2i}-\frac{2}{\tau_i}\boldsymbol{F}_{3i})+\frac{5}{\tau_i}(\boldsymbol{F}_{2i}-\frac{6}{\tau_i}\boldsymbol{F}_{3i}+\frac{12}{\tau_i^2}\boldsymbol{F}_{4i})^{\top}\boldsymbol{R}_{1i}(\boldsymbol{F}_{2i}-\frac{6}{\tau_i}\boldsymbol{F}_{3i}+\frac{12}{\tau_i^2}\boldsymbol{F}_{4i}),\ i=1,\cdots,w$
$\boldsymbol{\Psi}_i=\frac{2}{\tau_i^2}\boldsymbol{F}_{3i}^{\top}\boldsymbol{R}_{2i}\boldsymbol{F}_{3i}+\frac{16}{\tau_i^2}(\boldsymbol{F}_{3i}-\frac{3}{\tau_i}\boldsymbol{F}_{4i})^{\top}\boldsymbol{R}_{2i}(\boldsymbol{F}_{3i}-\frac{3}{\tau_i}\boldsymbol{F}_{4i}),\ i=1,\cdots,w$
$\boldsymbol{\eta}_i(t)=\mathrm{col}\left\{\boldsymbol{x}(t),\int_{t-\tau_i}^{t}\mathrm{e}^{\epsilon(s-t+\tau_i)}\boldsymbol{x}_r(s)ds,\int_{-\tau_i}^{0}\int_{t+\theta}^{t}\mathrm{e}^{\epsilon(s-t+\tau_i)}\boldsymbol{x}_r(s)dsd\theta,\int_{-\tau_i}^{0}\int_{\beta}^{0}\int_{t+\theta}^{t}\mathrm{e}^{\epsilon(s-t+\tau_i)}\boldsymbol{x}_r(s)dsd\theta d\beta\right\}$
$\boldsymbol{\varphi}_i(t)=\mathrm{col}\left\{\boldsymbol{x}_r(t-\tau_i),\int_{t-\tau_i}^{t}\mathrm{e}^{\epsilon(s-t+\tau_i)}\boldsymbol{x}_r(s)ds,\int_{-\tau_i}^{0}\int_{t+\theta}^{t}\mathrm{e}^{\epsilon(s-t+\tau_i)}\boldsymbol{x}_r(s)dsd\theta,\int_{-\tau_i}^{0}\int_{\beta}^{0}\int_{t+\theta}^{t}\mathrm{e}^{\epsilon(s-t+\tau_i)}\boldsymbol{x}_r(s)dsd\theta d\beta\right\}$
$\boldsymbol{\varphi}(t)=\mathrm{col}\left\{\boldsymbol{x}(t),\boldsymbol{\varphi}_1(t),\cdots,\boldsymbol{\varphi}_w(t)\right\}$

$$\boldsymbol{C}_{\tau i}=\boldsymbol{c}_{ri}d_{ci}\boldsymbol{e}_i^{\top},\ \ \boldsymbol{D}=\sum_{i=1}^{w+l}\boldsymbol{e}_i d_{ci}\boldsymbol{e}_i^{\top} \tag{9-11}$$

where $\boldsymbol{e}_i$ is the unit vector such that $u_i=\boldsymbol{e}_i^{\top}\boldsymbol{u}$.

9.3 Stability and Dynamical Performance Analysis

This section presents the theory foundation of analysis and design on the time-delay reduced system Φ first without then with consideration of the model reduction error system Δ. All proofs of the theorems are included in the support materials and all the special symbols are defined in Table I.

9.3.1 Stability and Performance of the Reduced-order System

We begin with establishing the ϵ-exponential stability and H_2 performance analysis results of the system Φ without considering the system Δ. In power system damping control, damping factor, the negative real part of the rightmost eigenvalue, characterizes the rate of disturbance suppression and it is strongly related to the settling time of oscillations. In addition, it underlays the concept of *damping factor based delay margin* for power system damping control [54, 55]. From a time-domain viewpoint, damping factor is equivalent to the concept of ϵ-exponential stability. A time-delay system is said to be ϵ-exponentially stable if there exist scalar ϵ and $\alpha>0$ such that $\|\boldsymbol{x}(t)\|\leq\alpha e^{-\epsilon t}\|\boldsymbol{\phi}\|_d$ where $\boldsymbol{\phi}\in C([-\tau_m,0],\mathbb{R}^n)$ is the initial state of the time-delay system and $\|\boldsymbol{\phi}\|_d=\sup_{-\tau_m\leq\theta\leq 0}\|\boldsymbol{\phi}(\theta)\|$. The following result establishes the ϵ-exponential stability of the system Φ.

Theorem 9.1： System Φ is ϵ-exponentially stable if there exist symmetric positive definite matrices $\boldsymbol{P}_i\in\mathbb{R}^{(4n+m)\times(4n+m)}$, $\boldsymbol{R}_{0i},\boldsymbol{R}_{1i},\boldsymbol{R}_{2i}\in\mathbb{R}^{n\times n}$, $i=1,\cdots,w$, such that $\boldsymbol{\Theta}(\epsilon)\leq 0$. □

Proof： Construct a candidate LKF as follows

$$V_\Phi=\sum_{i=1}^{w}V_{0i}+W_{0i}+W_{1i}+W_{2i} \tag{9-12}$$

where

$$V_{0i} = \boldsymbol{\eta}_i^\top(t)\boldsymbol{P}_i\boldsymbol{\eta}_i(t) \tag{9-13}$$

$$W_{0i} = \int_{t-\tau_i}^{t} \mathrm{e}^{2\epsilon(s-t+\tau_i)}\boldsymbol{x}_r^\top(s)\boldsymbol{R}_{0i}\boldsymbol{x}_r(s)ds \tag{9-14}$$

$$W_{1i} = \int_{-\tau_i}^{0}\int_{t+\theta}^{t} \mathrm{e}^{2\epsilon(s-t+\tau_i)}\boldsymbol{x}_r^\top(s)\boldsymbol{R}_{1i}\boldsymbol{x}_r(s)dsd\theta \tag{9-15}$$

$$W_{2i} = \int_{-\tau_i}^{0}\int_{\beta}^{0}\int_{t+\theta}^{t} \mathrm{e}^{2\epsilon(s-t+\tau_i)}\boldsymbol{x}_r^\top(s)\boldsymbol{R}_{2i}\boldsymbol{x}_r(s)dsd\theta d\beta. \tag{9-16}$$

Calculating the derivatives along the trajectory of system $\boldsymbol{\Phi}$, we have

$$\dot{V}_{0i} + 2\epsilon V_{0i} = \boldsymbol{\varphi}(t)^\top(\boldsymbol{K}_i^\top\boldsymbol{P}_i\boldsymbol{J}_i + \boldsymbol{J}_i\boldsymbol{P}_i\boldsymbol{K}_i + 2\epsilon\boldsymbol{J}_i^\top\boldsymbol{P}_i\boldsymbol{J}_i)\boldsymbol{\varphi}(t) \tag{9-17}$$

$$\dot{W}_{0i} + 2\epsilon W_{0i} = \boldsymbol{\varphi}^\top(t)(\mathrm{e}^{2\epsilon\tau_i}\boldsymbol{E}_{0r}^\top\boldsymbol{R}_{0i}\boldsymbol{E}_{0r} - \boldsymbol{F}_{1i}^\top\boldsymbol{R}_{0i}\boldsymbol{F}_{1i})\boldsymbol{\varphi}(t) \tag{9-18}$$

$$\begin{aligned}\dot{W}_{1i} + 2\epsilon W_{1i} =&\boldsymbol{\varphi}^\top(t)(\tau_i\mathrm{e}^{2\epsilon\tau_i}\boldsymbol{E}_{0r}^\top\boldsymbol{R}_{1i}\boldsymbol{E}_{0r})\boldsymbol{\varphi}(t)\\ &- \int_{t-\tau_i}^{t} \mathrm{e}^{2\epsilon(s-t+\tau_i)}\boldsymbol{x}_r^\top(s)\boldsymbol{R}_{1i}\boldsymbol{x}_r(s)ds\end{aligned} \tag{9-19}$$

$$\begin{aligned}\dot{W}_{2i} + 2\epsilon W_{2i} =&\boldsymbol{\varphi}^\top(t)(\frac{\tau_i^2}{2}\mathrm{e}^{2\epsilon\tau_i}\boldsymbol{E}_{0r}^\top\boldsymbol{R}_{2i}\boldsymbol{E}_{0r})\boldsymbol{\varphi}(t)\\ &- \int_{-\tau_i}^{0}\int_{t+\theta}^{t} \mathrm{e}^{2\epsilon(s-t+\tau_i)}\boldsymbol{x}_r^\top(s)\boldsymbol{R}_{2i}\boldsymbol{x}_r(s)dsd\theta.\end{aligned} \tag{9-20}$$

Using Lemma 3 and Lemma 4 in [204] to estimate the integral terms in (9-19) and (9-20) yields

$$\int_{t-\tau_i}^{t} \mathrm{e}^{2\epsilon(s-t+\tau_i)}\boldsymbol{x}_r^\top(s)\boldsymbol{R}_{1i}\boldsymbol{x}_r(s)ds \geq \boldsymbol{\varphi}^\top(t)\boldsymbol{\Pi}_i\boldsymbol{\varphi}(t) \tag{9-21}$$

$$\int_{-\tau_i}^{0}\int_{t+\theta}^{t} \mathrm{e}^{2\epsilon(s-t+\tau_i)}\boldsymbol{x}_r^\top(s)\boldsymbol{R}_{2i}\boldsymbol{x}_r(s)dsd\theta \geq \boldsymbol{\varphi}^\top(t)\boldsymbol{\Psi}_i\boldsymbol{\varphi}(t). \tag{9-22}$$

Substituting (9-17), (9-18), (9-19) and (9-20) into (9-12), we prove that $\boldsymbol{\Theta}(\epsilon) \preccurlyeq 0$ leads to $\dot{V}_\Phi + 2\epsilon V_\Phi \leq 0$. By Gronwall Lemma [205, Lemma 2.2], $V_\Phi(t) \leq e^{-2\epsilon t}V_\Phi(0)$. Due to the positive definiteness of all weighting matrices, it is trivial that there exist scalar $\alpha, \beta > 0$ such that $\alpha\|\boldsymbol{\phi}\|_d^2 \leq V_\Phi(\boldsymbol{\phi}) \leq \beta\|\boldsymbol{\phi}\|_d^2$ where $\boldsymbol{\phi}(\theta) = \boldsymbol{x}_r(\theta), \theta \in [-\tau_m, 0]$ is the initial condition of sysmtem Φ. Therefore, $\|\boldsymbol{x}_r(t)\| \leq e^{-\epsilon t}\sqrt{\frac{\beta}{\alpha}}\|\boldsymbol{\phi}\|_d, \forall t \geq 0$. ■

Merely a satisfactory damping factor is not enough to ensure a good dynamical performance. Another performance metric is needed. The most widely adopted dynamical perfor-

mance metric in power systems is the damping ratio of electromechanical modes. Maintaining a minimum damping ratio for the critical modes is a well-accepted design specification for damping control systems [202]. However, the damping ratio, as a frequency domain metric, is quite difficult to be considered in a time-domain design framework. To address this issue, we consider the following H_2 performance metric

$$\mathcal{M}_r = \int_0^\infty \boldsymbol{x}_r^\top(t)\boldsymbol{Q}x_r(t)dt \tag{9-23}$$

where the weighting matrice $\boldsymbol{Q}$ is carefully selected as

$$\boldsymbol{Q} = \sum_{i=1}^{2m} \frac{\xi_*\Omega_i}{\sqrt{1-\xi_*^2}}\boldsymbol{v}_i\boldsymbol{v}_i^\top \tag{9-24}$$

where Ω_i is the imaginary part of the ith critical mode of $\boldsymbol{A}_r$ and $\boldsymbol{v}_i$ is the corresponding left eigenvector, i.e. the observable direction of the mode; ξ_* is the required damping ratio for all critical modes. The coefficient $\xi_*\Omega_i/\sqrt{1-\xi_*^2}$ denotes the minimum negative real part needed for the ith mode to have the required damping ratio ξ_*. Because the output energy of a mode is inversely proportional to the negative real part, the designed coefficients make the output energy of each critical modes on approximately the same scale. Then minimizing the metric $\mathcal{M}_r$ could push each critical mode leftward.

Assume that some instantaneous disturbance happens at $t = 0$ and it brings the states to $\boldsymbol{x}_r(0) \neq \boldsymbol{0}$. In this case, the following result establishes a upper bound of the performance metric $\mathcal{M}_r$ for any disturbance with $\boldsymbol{x}_r(0)^\top\boldsymbol{x}_r(0) \leq 1$.

Theorem 9.2： The system Φ is Lyapunov stable and the H_2 performance metric $\mathcal{M}_r \leq \sigma$ for any $\boldsymbol{x}_r(0)^\top\boldsymbol{x}_r(0) \leq 1$ if there exist symmetric positive definite matrices $\boldsymbol{P}_i \in \mathbb{R}^{(4n+m)\times(4n+m)}$, $\boldsymbol{R}_{0i}, \boldsymbol{R}_{1i}, \boldsymbol{R}_{2i} \in \mathbb{R}^{n\times n}$, $i = 1, \cdots, w$ and $\eta > 0$ such that

$$\boldsymbol{\Theta}(0) + \eta\boldsymbol{Y} \preccurlyeq 0 \tag{9-25}$$

$$\sum_{i=1}^{w} \boldsymbol{N}_i\boldsymbol{P}_i\boldsymbol{N}_i^\top \prec \sigma\eta\boldsymbol{I} \tag{9-26}$$

where $\boldsymbol{Y} = \boldsymbol{E}_{0r}^\top\boldsymbol{Q}\boldsymbol{E}_{0r}$. □

Proof： Following the proof of theorem 9.1, matrix inequality (9-26) implies

$$-\eta\boldsymbol{x}_r^\top\boldsymbol{E}_{0r}^\top\boldsymbol{Q}\boldsymbol{E}_{0r}\boldsymbol{x}_r \geq \dot{V}_\Phi \tag{9-27}$$

Therefore, full-order close-loop system Π is Lyapunov stable. Integraring (9-27) yields

$$V_\Phi(0) \geq \eta\int_0^\infty \boldsymbol{x}_r^\top\boldsymbol{E}_{0r}^\top\boldsymbol{Q}\boldsymbol{E}_{0r}\boldsymbol{x}_r = \eta\mathcal{M}_r \tag{9-28}$$

and LMI (9-26) ensures $\sigma\eta\boldsymbol{x}_r(0)^\top\boldsymbol{x}_r(0) > V_\Phi(0)$ which implies $\mathcal{M}_r < \sigma, \forall\boldsymbol{x}_r(0) \in \mathbb{R}^n$ with $\boldsymbol{x}_r(0)^\top\boldsymbol{x}_r(0) \leq 1$. ■

Note that the introduction of scalar η is theoretically unnecessary, but it significantly improves the numerical stability of the design algorithm to be presented section 9.4. In the algorithm, η is treated as a flexible decision variable which alleviates the ill-conditioning of the SDP problems.

9.3.2 Considering the Impact of Error System in Stability Analysis

The stability condition in theorem 9.1 completely ignores the information of the reduction error system Δ. It purely focuses on reduced system Φ and says nothing about the stability of the original full-order system. As a result, the stability and damping factor of the full-order system are still in doubt. In this subsection, we propose a method to ensure the ϵ-exponential stability of the full-order system by just using the model of the reduced system Φ and H_∞ error of the model reduction.

The concept of dissipativity [205] is the foundation:

Definition 9.1： Consider a dynamical system with input $\boldsymbol{u}$, state $\boldsymbol{x}$ and output $\boldsymbol{y}$. The system is *exponentially dissipative* with the supply rate $r(\boldsymbol{u}, \boldsymbol{y})$ and decay rate ϵ if there exists a non-negative function $V(\boldsymbol{x})$, called storage function, such that

$$\dot{V}(\boldsymbol{x}) + 2\epsilon V(\boldsymbol{x}) \leq r(\boldsymbol{u}(t), \boldsymbol{y}(t)) \tag{9-29}$$

along the trajectory of the system where $r(\boldsymbol{u}, \boldsymbol{y})$ is integrable in any finite time interval and $V(\boldsymbol{x})$ is continuous and $r(\mathbf{0}, \mathbf{0}) = 0$ and $V(\mathbf{0}) = 0$ at the steady state. ◊

The following result gives sufficient condition for the ϵ-exponential stability of the full-order system Π by viewing Π as the feedback interconnect of reduced system Φ and error system Δ.

Theorem 9.3： Asume that the error system $\boldsymbol{\Delta}$ and the reduced close-loop system Φ satisfy:

1. system Δ is exponentially dissipative with supply rate $\gamma^2\|\boldsymbol{\omega}\|^2 - \|\boldsymbol{z}\|^2$ and decay rate ϵ. For the associated storage function $V_\Delta(\boldsymbol{\xi}_t)$, there exist positive scalars α', β' such that $\alpha'\|\boldsymbol{\xi}(t)\|^2 \leq V_\Delta(\boldsymbol{\xi}_t) \leq \beta'\|\boldsymbol{\xi}_t\|_d^2$ where $\boldsymbol{\xi}_t$ is the state function of time-delay system Δ;
2. system Φ is exponentially dissipative with supply rate $\frac{1}{\gamma^2}\|\boldsymbol{u}\|^2 - \|\boldsymbol{y}\|^2$ and decay rate ϵ. For the associated storage function $V_\Phi(\boldsymbol{x}_t)$, there exist positive scalars α'', β'' such that $\alpha''\|\boldsymbol{x}(t)\|^2 \leq V_\Phi(\boldsymbol{x}_t) \leq \beta''\|\boldsymbol{x}_t\|_d^2$ where $\boldsymbol{x}_t$ is the state function of time-delay system Φ;

Then the full-order close-loop system Π is exponentially stable with decay rate ϵ. □

Proof： Condition 1) and 2) imply that there exist non-negative function $V_\Delta(\boldsymbol{\xi})$ and func-

tional $V_\Phi(\boldsymbol{x}_t)$ such that

$$\begin{cases} \dot{V}_\Delta(t) + 2\epsilon V_\Delta(t) \le \gamma^2\|\boldsymbol{\omega}\|^2 - \|\boldsymbol{z}\|^2 \\ \dot{V}_\Phi(t) + 2\epsilon V_\Phi(t) \le \frac{1}{\gamma^2}\|\boldsymbol{u}\|^2 - \|\boldsymbol{y}\|^2. \end{cases} \tag{9-30}$$

For feedback interconnection, we have $\boldsymbol{u} = \boldsymbol{z}$ and $\boldsymbol{y} = \boldsymbol{\omega}$. Consider a candidate Lyapunov functional $V(t) = V_\Delta(t) + \gamma^2 V_\Phi(t)$ for full-order close-loop system Π. It follows from (9-30) that

$$\begin{aligned} \dot{V}(t) &= \dot{V}_\Delta(t) + \gamma^2\dot{V}_\Phi(t) \\ &\le -2\epsilon(V_\Delta(t) + \gamma^2 V_\Phi(t) \\ &\le -2\epsilon V(t) \end{aligned} \tag{9-31}$$

which gives $V(t) \le e^{-2\epsilon t}V(0)$, $t \ge 0$ by Gronwall Lemma [205, Lemma 2.2], $V(t) \le e^{-2\epsilon t}V(0)$. Furthermore, taking $\alpha = \min\{\alpha', \gamma^2\alpha''\}$ and $\beta = \max\{\beta', \gamma^2\beta''\}$, we have

$$\begin{aligned} \alpha\|(\boldsymbol{\xi}(t), \boldsymbol{x}(t))\|^2 &\le \alpha'\|\boldsymbol{\xi}(t)\|^2 + \gamma^2\alpha''\|\boldsymbol{x}(t)\|^2 \le V(t) \\ &\le e^{-2\epsilon t}V(0) \le \beta'\|\boldsymbol{\xi}_0\|_d^2 + \gamma^2\beta''\|\boldsymbol{x}_0\|_d^2 \le \beta\|(\boldsymbol{\xi}_0, \boldsymbol{x}_0)\|_d^2 \end{aligned} \tag{9-32}$$

which yields $\|(\boldsymbol{\xi}(t), \boldsymbol{x}(t))\| \le e^{-\epsilon t}\sqrt{\frac{\beta}{\alpha}}\|(\boldsymbol{\xi}_0, \boldsymbol{x}_0)\|$. ■

The remaining tasks are to make sure system Δ and Φ respectively satisfy condition 1) and condition 2) of above theorem. We first propose the following procedure, named *dissipative model reduction (DMR)*, to obtain a reduced-order model Σ_r for the full-order power system Σ_s such that the error system Δ satisfies condition 1) of theorem 9.3.

Step 1: Find a state transformation $\boldsymbol{T}$ of Σ_s such that $\boldsymbol{T}\boldsymbol{A}_s\boldsymbol{T}^{-1} = diag(\boldsymbol{A}_s', \boldsymbol{A}_s'')$

$$\left[\begin{array}{c|c} \boldsymbol{T}\boldsymbol{A}_s\boldsymbol{T}^{-1} & \boldsymbol{T}\boldsymbol{B}_s \\ \hline \boldsymbol{C}_s\boldsymbol{T}^{-1} & \end{array}\right] = \left[\begin{array}{cc|c} \boldsymbol{A}_s' & & \boldsymbol{B}_s' \\ & \boldsymbol{A}_s'' & \boldsymbol{B}_s'' \\ \hline \boldsymbol{C}_s' & \boldsymbol{C}_s'' & \end{array}\right] \tag{9-33}$$

where $\mathrm{Re}\lambda(\boldsymbol{A}_s') < -\epsilon$ and $\boldsymbol{A}_s''$ contains all eigenvalues lying on the right of $-\epsilon$. Other critical modes can also be included in $\boldsymbol{A}_s''$ in case that they get lost in the model reduction. Such $\boldsymbol{T}$ can be obtained by performing an ordered Schur factorization followed by solving a Sylvester equation similar to the approach detailed in the appendix of [206].

Step 2: Apply any stability-preserving model reduction approach, e.g. balanced truncation, Hankel-norm approximation etc. [207], to the system $\Sigma_s'(\boldsymbol{A}_s' + \epsilon\boldsymbol{I}, \boldsymbol{B}_s', \boldsymbol{C}_s')$ and obtain a reduced system $\Sigma_r'(\boldsymbol{A}_r', \boldsymbol{B}_r', \boldsymbol{C}_r')$ with a H_∞ error bound $\|\Sigma_r' - \Sigma_s'\|_\infty \le \gamma_0$.

Step 3: Then we construct the reduced model Σ_r as

$$\left[\begin{array}{c|c} \boldsymbol{A}_r & \boldsymbol{B}_r \\ \hline \boldsymbol{C}_r & \end{array}\right] = \left[\begin{array}{cc|c} \boldsymbol{A}_r' - \epsilon\boldsymbol{I} & & \boldsymbol{B}_r' \\ & \boldsymbol{A}_s'' & \boldsymbol{B}_s'' \\ \hline \boldsymbol{C}_r' & \boldsymbol{C}_s'' & \end{array}\right]. \tag{9-34}$$

The following result ensures the validity and usefulness of the DMR procedure.

Theorem 9.4： The delayed error system Δ shown in Fig. 9-2 obtained from the DMR procedure satisfies condition 1) of theorem 9.1 with $\gamma = \gamma_0 \mathrm{e}^{\epsilon\tau_{max}}$. □

Proof： We first condier the error system $\Sigma_s - \Sigma_r$. From Step 1 of DMR procedure, the system $\Sigma_s'(\boldsymbol{A}_s' + \epsilon\boldsymbol{I}, \boldsymbol{B}_s', \boldsymbol{C}_s')$ is stable since $\mathrm{Re}\lambda(\boldsymbol{A}_s') < -\epsilon$. According to Step 2, the error system

$$\Sigma_r' - \Sigma_s' = \left[\begin{array}{cc|c} \boldsymbol{A}_r' & & \boldsymbol{B}_r' \\ & \boldsymbol{A}_s' + \epsilon\boldsymbol{I} & \boldsymbol{B}_s' \\ \hline \boldsymbol{C}_r' & -\boldsymbol{C}_s' & \end{array}\right] \tag{9-35}$$

is also stable because $\boldsymbol{A}_r'$ preserves the stability of $\boldsymbol{A}_s' + \epsilon\boldsymbol{I}$. Furthermore, $\|\Sigma_r' - \Sigma_s'\|_\infty \leq \gamma_0$. For brevity, we define $\boldsymbol{A}_\Delta = \mathrm{diag}(\boldsymbol{A}_r' - \epsilon\boldsymbol{I}, \boldsymbol{A}_s')$, $\boldsymbol{B}_\Delta = [\boldsymbol{B}_r'^\top \ \boldsymbol{B}_s'^\top]^\top$ and $\boldsymbol{C}_\Delta = [\boldsymbol{C}_r' \ \boldsymbol{C}_s']$. Thus, by the bounded real lemma [208, Lemma 8.1], there exists a positive definite matrix $\boldsymbol{P}$ such that

$$\begin{bmatrix} (\boldsymbol{A}_\Delta + \epsilon\boldsymbol{I})^\top\boldsymbol{P} + \boldsymbol{P}(\boldsymbol{A}_\Delta + \epsilon\boldsymbol{I}) + \boldsymbol{C}_\Delta^\top\boldsymbol{C}_\Delta & \boldsymbol{P}\boldsymbol{B}_\Delta \\ \boldsymbol{B}_\Delta^\top\boldsymbol{P} & -\gamma_0^2\boldsymbol{I} \end{bmatrix} \prec 0 \tag{9-36}$$

In addition, following Step 3, the error system

$$\Delta = \left[\begin{array}{cc|c} \boldsymbol{A}_r' - \epsilon\boldsymbol{I} & & \boldsymbol{B}_r' \\ & \boldsymbol{A}_s' & \boldsymbol{B}_s' \\ \hline \boldsymbol{C}_r' & -\boldsymbol{C}_s' & \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{A}_\Delta & \boldsymbol{B}_\Delta \\ \hline \boldsymbol{C}_\Delta & \end{array}\right]. \tag{9-37}$$

Therefore, given a candidate storage function $V_1(\boldsymbol{\xi}) = \boldsymbol{\xi}^\top\boldsymbol{P}\boldsymbol{\xi}$, the matrix inequality (9-36) is equivalent to

$$\dot{V}_1(\boldsymbol{\xi}) + 2\epsilon V_1(\boldsymbol{\xi}) \leq \gamma_0^2\|\boldsymbol{\omega}\|^2 - \|\boldsymbol{v}\|^2. \tag{9-38}$$

Moreover, $\underline{\sigma}\|\boldsymbol{\xi}\|^2 \leq V_1(\boldsymbol{\xi}) \leq \overline{\sigma}\|\boldsymbol{\xi}\|^2$ where $\underline{\sigma}$ and $\overline{\sigma}$ are the smallest and largest eigenvalues of $\boldsymbol{P}$, respectively. Due to the positive definitness of $\boldsymbol{P}$, $\overline{\sigma} > \underline{\sigma} > 0$.

We then consider the delay block, i.e.

$$\boldsymbol{z} = \sum_{i=1}^{w} \boldsymbol{e}_i v_i(t - \tau_i) + \sum_{i=w+1}^{w+l} \boldsymbol{e}_i v_i(t). \tag{9-39}$$

Consider the candidate storage function

$$V_2 = \sum_{i=1}^{w} \int_{t-\tau_i}^{t} \mathrm{e}^{2\epsilon(s-t+\tau_i)} v_i^2(s) ds. \tag{9-40}$$

We have

$$\begin{aligned}\dot{V}_2 + 2\epsilon V_2 &= \sum_{i=1}^{w} \mathrm{e}^{2\epsilon\tau_i} v_i^2(t) - v_i^2(t-\tau_i) \\ &\le \sum_{i=1}^{w} \mathrm{e}^{2\epsilon\tau_m} v_i^2(t) - v_i^2(t-\tau_i) \\ &= \sum_{i=1}^{w} \mathrm{e}^{2\epsilon\tau_m} v_i^2(t) - z_i^2(t) \\ &\le \mathrm{e}^{2\epsilon\tau_m} \|\boldsymbol{v}\|^2 - \|\boldsymbol{z}\|^2.\end{aligned} \tag{9-41}$$

Multiplying (9-38) by $\mathrm{e}^{2\epsilon\tau_m}$ and adding it to (9-41) yield

$$\dot{V}_\Delta + 2\epsilon V_\Delta \le \gamma_0^2 \mathrm{e}^{2\epsilon\tau_m} \|\boldsymbol{\omega}\|^2 - \|\boldsymbol{z}\|^2 \tag{9-42}$$

where $V_\Delta = \mathrm{e}^{2\epsilon\tau_m} V_1 + V_2$ which is considered as the storage function for the system Δ. ■

The exponential dissipativity of reduced system Φ is established in the following result.

Theorem 9.5： System $\boldsymbol{\Phi}$ given by (9-8) and (9-9) satisfies condition 2) of theorem 9.2 if there exist symmetric positive definite matrices $\boldsymbol{P}_i \in \mathbb{R}^{(4n+m)\times(4n+m)}$, $\boldsymbol{R}_{0i}, \boldsymbol{R}_{1i}, \boldsymbol{R}_{2i} \in \mathbb{R}^{n\times n}$, $i = 1, \cdots, w$, such that the following matrix inquality holds

$$\begin{bmatrix} \boldsymbol{\Theta}(\epsilon) & \sum_{i=1}^{w} \boldsymbol{J}_i^\top \boldsymbol{P}_i \boldsymbol{B}_\tau & \boldsymbol{C}_\tau \\ \boldsymbol{B}_\tau^\top \sum_{i=1}^{w} \boldsymbol{P}_i \boldsymbol{J}_i & -\frac{1}{\gamma}\boldsymbol{I} & \boldsymbol{D}^\top \\ \boldsymbol{C}_\tau^\top & \boldsymbol{D} & -\frac{1}{\gamma}\boldsymbol{I} \end{bmatrix} \prec 0 \tag{9-43}$$

□

Proof： Construct the same LKF as in theorem 9.1 and estimate the derivatives in the same way. The only difference is

$$\dot{V}_{0i} + 2\epsilon V_{0i} = \begin{bmatrix} \boldsymbol{\varphi}(t) \\ \boldsymbol{u}(t) \end{bmatrix}^\top \begin{bmatrix} \boldsymbol{K}_i^\top \boldsymbol{P}_i \boldsymbol{J}_i + \boldsymbol{J}_i \boldsymbol{P}_i \boldsymbol{K}_i + 2\epsilon \boldsymbol{J}_i^\top \boldsymbol{P}_i \boldsymbol{J}_i & \boldsymbol{J}_i^\top \boldsymbol{P}_i \boldsymbol{B}_\tau \\ \boldsymbol{B}_\tau^\top \boldsymbol{P}_i \boldsymbol{J}_i & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varphi}(t) \\ \boldsymbol{u}(t) \end{bmatrix} \tag{9-44}$$

i.e. the input of system Φ from the error system output needs to be considered. Further note from (9-9) that

$$\frac{1}{\gamma^2}\boldsymbol{u}^\top \boldsymbol{u} - \boldsymbol{y}^\top \boldsymbol{y} = \begin{bmatrix} \boldsymbol{\varphi}(t) \\ \boldsymbol{u}(t) \end{bmatrix}^\top \begin{bmatrix} -\boldsymbol{C}_\tau \boldsymbol{C}_\tau^\top & -\boldsymbol{C}_\tau \boldsymbol{D} \\ -\boldsymbol{D}^\top \boldsymbol{C}_\tau & \frac{1}{\gamma^2}\boldsymbol{I} - \boldsymbol{D}^\top \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \boldsymbol{\varphi}(t) \\ \boldsymbol{u}(t) \end{bmatrix} \tag{9-45}$$

Substituting (9-44), (9-18), (9-19) and (9-20) into (9-12) and further considering (9-45), we prove that

$$\begin{bmatrix} \boldsymbol{\Theta}(\epsilon) + \boldsymbol{C}_\tau \boldsymbol{C}_\tau^\top & \boldsymbol{C}_\tau \boldsymbol{D} + \sum_{i=1}^{w} \boldsymbol{J}_i^\top \boldsymbol{P}_i \boldsymbol{B}_\tau \\ \boldsymbol{D}^\top \boldsymbol{C}_\tau^\top + \boldsymbol{B}_\tau^\top \sum_{i=1}^{w} \boldsymbol{P}_i \boldsymbol{J}_i & -\frac{1}{\gamma^2}\boldsymbol{I} + \boldsymbol{D}^\top \boldsymbol{D} \end{bmatrix} \prec 0 \tag{9-46}$$

implies $\dot{V}_\Phi + 2\epsilon V_\Phi \le \frac{1}{\gamma^2}\boldsymbol{u}^\top \boldsymbol{u} - \boldsymbol{y}^\top \boldsymbol{y}$. Multiplying both side of (9-46) with γ, renaming $\gamma\boldsymbol{P}_i, \gamma\boldsymbol{R}_{0i}, \gamma\boldsymbol{R}_{1i}, \gamma\boldsymbol{R}_{2i}$ as $\boldsymbol{P}_i, \boldsymbol{R}_{0i}, \boldsymbol{R}_{1i}, \boldsymbol{R}_{2i}$ and using Schur's lemma, (9-46) is equivalent to

(9-43). ■

In short, following the DMR procedure to obtain the reduced-order system and enforcing the ϵ-exponential dissipativity condition on the reduced system, together will guarantee that the original full-order system is ϵ-exponentially stable.

9.4 Control Design

9.4.1 Optimization Problem Formulation

To alleviate the notational burden, we write $\boldsymbol{S}_\tau = \{\boldsymbol{A}_\tau, \boldsymbol{B}_\tau, \boldsymbol{C}_\tau, \boldsymbol{D}\}$ to condense all system information needed in theorem 9.5 and 9.2 in a single symbol. Similarly, we denote $\boldsymbol{P} = \{\boldsymbol{P}_i\}_{1\le i\le w}$ and $\boldsymbol{R} = \{\boldsymbol{R}_{1i}, \boldsymbol{R}_{2i}, \boldsymbol{R}_{3i}\}_{1\le i\le w}$. We abuse the notation $\boldsymbol{P} > 0$ and $\boldsymbol{R} > 0$ to mean $\boldsymbol{P}_i > 0$ and $\boldsymbol{R}_{1i}, \boldsymbol{R}_{2i}, \boldsymbol{R}_{3i} > 0$, $\forall 1 \le i \le w$, respectively. Furthermore, we can abstract the matrix inequality (9-43) in a compact form as

$$\boldsymbol{H}_b(\epsilon, \boldsymbol{S}_\tau, \boldsymbol{P}) + \boldsymbol{H}_l(\epsilon, \boldsymbol{R}) \prec 0. \tag{9-47}$$

where $\boldsymbol{H}_b(\cdot)$ is the part of the matrix in (9-43) which depends biliearly on $\boldsymbol{S}_\tau$ and $\boldsymbol{P}$, i.e.

$$\boldsymbol{H}_b(\epsilon, \boldsymbol{S}_\tau, \boldsymbol{P}) = \begin{bmatrix} \sum_{i=1}^w \boldsymbol{A}_{\eta i}^\top \boldsymbol{P}_i \boldsymbol{J}_i + \boldsymbol{J}_i^\top \boldsymbol{P}_i \boldsymbol{A}_{\eta i} + 2\epsilon \boldsymbol{J}_i^\top \boldsymbol{P}_i \boldsymbol{J}_i & \sum_{i=1}^w \boldsymbol{J}_i^\top \boldsymbol{P}_i \boldsymbol{B}_\tau & \boldsymbol{0} \\ \boldsymbol{B}_\tau^\top \sum_{i=1}^w \boldsymbol{P}_i \boldsymbol{J}_i & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{9-48}$$

The rest part of the matrix in (9-43) contains only linear terms and is written as $\boldsymbol{H}_l(\cdot)$:

Similar, matrix inequality (9-25) is also break into a bilinear part and the rest linear part:

$$\boldsymbol{\Theta}_b(0, \boldsymbol{S}_\tau, \boldsymbol{P}) + \boldsymbol{\Theta}_l(0, \boldsymbol{R}) + \eta \boldsymbol{E}_{0r}^\top \boldsymbol{Q} \boldsymbol{E}_{0r} \prec 0. \tag{9-49}$$

We further write the LMI (9-26) in a compact form as

$$\boldsymbol{G}(\boldsymbol{P}) \prec \sigma\eta \boldsymbol{I}. \tag{9-50}$$

Suppose that the system description $\boldsymbol{S}_\tau$ depends on the controller parameters $\boldsymbol{p} \in \mathbb{R}^q$, i.e. $\boldsymbol{S}_\tau(\boldsymbol{p})$. The optimal control design problem here aims at obtaining the controller parameters which guarantee that close-loop system is exponentially stable with decay rate ϵ while the H_2 performance metric $\mathcal{M}_r$ under unit-energy disturbance is minimized. The problem is formally stated below

$$\min_{\sigma, \boldsymbol{p}, \boldsymbol{P}, \boldsymbol{R}, \bar{\boldsymbol{P}}, \bar{\boldsymbol{R}}} \ \sigma \tag{9-51}$$

subject to

$$\underline{\boldsymbol{p}} \le \boldsymbol{p} \le \overline{\boldsymbol{p}} \tag{9-52a}$$

$$\boldsymbol{H}_b(\epsilon, \boldsymbol{S}_\tau(\boldsymbol{p}), \boldsymbol{P}) + \boldsymbol{H}_l(\epsilon, \boldsymbol{R}) \prec 0 \tag{9-52b}$$

$$\boldsymbol{\Theta}_b(0, \boldsymbol{S}_\tau, \bar{\boldsymbol{P}}) + \boldsymbol{\Theta}_l(0, \bar{\boldsymbol{R}}) + \eta \boldsymbol{E}_{0r}^\top \boldsymbol{Q} \boldsymbol{E}_{0r} \prec 0 \tag{9-52c}$$

$$\boldsymbol{G}(\bar{\boldsymbol{P}}) \prec \sigma\eta \boldsymbol{I} \tag{9-52d}$$

$$\boldsymbol{P}, \boldsymbol{R}, \bar{\boldsymbol{P}}, \bar{\boldsymbol{R}} \succ 0, \eta > 0 \tag{9-52e}$$

9.4.2 Path-following Method

The optimization problem (9-51)(9-52) is a nonlinear non-convex semi-definite program which is hopelessly intractable to be solved to global optimality. Hence a more reasonable goal is to design an algorithm to search for a local optimum. Starting from any initial control parameters, the path-following method presented here systematically reaches to a local optimum by iteratively linearizing the problem and solving an SDP in each iteration. Our implementation of the path-following method is an extension of the path-following method in [209, 210] to a general nonlinear SDP setting. Our method consists of two phases. Phase 1 looks for the control parameters to guarantee ϵ-exponential stability of the system. Phase 2 then optimize the H_2 performance while maintaining the ϵ-exponential stability obtained in phase 1.

Phase 1: Starting from a initial value of controller parameters $\boldsymbol{p}^{(0)}$, the objective of this phase is to find the controller parameters such that constraint (9-52a), (9-52b) and (9-52e) are feasible, i.e. the controller guarantees exponential stability of the close-loop system with prescribed decay rate ϵ. The steps are as follows.

(1.1) Find a $\beta^{(0)}$ as large as possible such that the LMI

$$\begin{aligned} &\min_{\boldsymbol{P}, \boldsymbol{R}, \rho} \ \rho \\ &\text{s.t.} \begin{cases} \boldsymbol{H}_b(\beta^{(0)}, \boldsymbol{S}_\tau(\boldsymbol{p}^{(0)}), \boldsymbol{P}) + \boldsymbol{H}_l(\beta^{(0)}, \boldsymbol{R}) \prec 0 \\ \rho \boldsymbol{I} \succ \boldsymbol{P}_i \succ \boldsymbol{0}, \forall i = 1, \cdots, w \\ \boldsymbol{R} \succ 0 \end{cases} \end{aligned} \tag{9-53}$$

is feasible with solution denoted as $\boldsymbol{P}^{(0)}$. If $\beta^{(0)} \geq \epsilon$, the work is done and $\boldsymbol{p}^{(0)}$ is the desirable control parameter. Otherwise, set $\theta = 0.1$ and $\delta\beta = 0.005$ and go to step (1.2).

(1.2) Set $\beta^{(1)} = \beta^{(0)} + \delta\beta$ and solve the following LMI with decision variable $\boldsymbol{\alpha}, \eta, \delta\boldsymbol{P}, \boldsymbol{R}$:

$$\begin{cases} \underline{\boldsymbol{p}} \le \boldsymbol{p}^{(0)} + \boldsymbol{\alpha} \le \overline{\boldsymbol{p}} \\ \boldsymbol{H}_b(\beta^{(1)}, \boldsymbol{S}_\tau(\boldsymbol{p}^{(0)}), \boldsymbol{P}^{(0)}) + \boldsymbol{H}_b(\beta^{(1)}, \delta\boldsymbol{S}_\tau, \boldsymbol{P}^{(0)}) \\ +\boldsymbol{H}_b(\beta^{(1)}, \boldsymbol{S}_\tau(\boldsymbol{p}^{(0)}), \delta\boldsymbol{P}) + \boldsymbol{H}_l(\beta^{(1)}, \boldsymbol{R}) \prec 0 \\ \boldsymbol{P}^{(0)} + \delta\boldsymbol{P} \succ 0, \boldsymbol{R} \succ 0 \\ \delta\boldsymbol{S}_\tau = \sum_{k=1}^{q} \frac{\partial \boldsymbol{S}_\tau}{\partial p_i}\Big|_{\boldsymbol{p}^{(0)}} \alpha_k \\ -\theta|\boldsymbol{p}^{(0)}| \le \boldsymbol{\alpha} \le \theta|\boldsymbol{p}^{(0)}| \\ \begin{bmatrix} 0.1\boldsymbol{P}_i^{(0)} & \delta\boldsymbol{P}_i \\ \delta\boldsymbol{P}_i & 0.1\boldsymbol{P}_i^{(0)} \end{bmatrix} \succ 0, \forall i = 1, \cdots, w \end{cases} \tag{9-54}$$

If it is feasible with solution $\boldsymbol{\alpha}$ and $\delta\boldsymbol{P}$, set $\boldsymbol{p}^{(1)} = \boldsymbol{p}^{(0)} + \boldsymbol{\alpha}$, $\boldsymbol{P}^{(1)} = \boldsymbol{P}^{(0)} + \delta\boldsymbol{P}$, and go to step (1.3). Otherwise, set $\delta\beta = \delta\beta/2$ and repeat step 1.2).

(1.3) Solve the SDP (9-53) with $\beta^{(0)}$ and $\boldsymbol{p}^{(0)}$ replaced with $\beta^{(1)}$ and $\boldsymbol{p}^{(1)}$. If the problem is feasible, set $\beta^{(0)} = \beta^{(1)}$, $\boldsymbol{p}^{(0)} = \boldsymbol{p}^{(1)}$ and $\boldsymbol{P}^{(0)} = \boldsymbol{P}^{(1)}$. Otherwise, set $\theta = \theta/2$ and go back to step (1.2). If $\beta^{(0)} \ge \epsilon$, the work is done and $\boldsymbol{p}^{(0)}$ is the desirable control parameter, otherwise go to step (1.2).

Phase 2: Optimize the performance.

(2.1) Find a $\sigma^{(0)}$ as large as possible such that the SDP

$$\min_{\boldsymbol{P}, \boldsymbol{R}, \bar{\boldsymbol{P}}, \bar{\boldsymbol{R}}, \eta, \rho} \quad \rho$$
$$\text{s.t.} \begin{cases} \boldsymbol{H}_b(\epsilon, \boldsymbol{S}_\tau(\boldsymbol{p}^{(0)}), \boldsymbol{P}) + \boldsymbol{H}_l(\epsilon, \boldsymbol{R}) \prec 0 \\ \boldsymbol{\Theta}_b(0, \boldsymbol{S}_\tau(\boldsymbol{p}^{(0)}), \bar{\boldsymbol{P}}) + \boldsymbol{\Theta}_l(0, \bar{\boldsymbol{R}}) + \eta\boldsymbol{Y} \prec 0 \\ \boldsymbol{G}(\bar{\boldsymbol{P}}) \prec \sigma^{(0)}\eta\boldsymbol{I} \\ \rho\boldsymbol{I} \succ \boldsymbol{P}_i \succ \boldsymbol{0}, \rho\boldsymbol{I} \succ \bar{\boldsymbol{P}}_i \succ \boldsymbol{0}, \forall i = 1, \cdots, w \\ \boldsymbol{R}, \bar{\boldsymbol{R}} \succ 0, \eta > 0 \end{cases} \tag{9-55}$$

is feasible with solution denoted as $\boldsymbol{P}^{(0)}$ and $\bar{\boldsymbol{P}}^{(0)}$. Set $\theta = 0.1$ and $\delta\sigma = 0.02\sigma^{(0)}$ and go to step (2.2).

(2.2) If $\delta\sigma$ is less than a threshold value or the maximal iteration number reached, stop and output the control parameter $\boldsymbol{p}^{(0)}$ with control performance $\sigma^{(0)}$. Otherwise set $\sigma^{(1)} = \sigma^{(0)} - \delta\sigma$. Solve the following LMI with decision variable $\boldsymbol{\alpha}, \eta, \delta\boldsymbol{P}, \delta\bar{\boldsymbol{P}}, \boldsymbol{R}, \bar{\boldsymbol{R}}$

$$
\begin{cases}
\underline{\boldsymbol{p}} \le \boldsymbol{p}^{(0)} + \boldsymbol{\alpha} \le \overline{\boldsymbol{p}} \\
\boldsymbol{H}_b(\epsilon, \boldsymbol{S}_\tau(\boldsymbol{p}^{(0)}), \boldsymbol{P}^{(0)}) + \boldsymbol{H}_b(\epsilon, \delta\boldsymbol{S}_\tau, \boldsymbol{P}^{(0)}) \\
+\boldsymbol{H}_b(\epsilon, \boldsymbol{S}_\tau(\boldsymbol{p}^{(0)}), \delta\boldsymbol{P}) + \boldsymbol{H}_l(\epsilon, \boldsymbol{R}) \prec 0 \\
\boldsymbol{\Theta}_b(0, \boldsymbol{S}_\tau(\boldsymbol{p}^{(0)}), \bar{\boldsymbol{P}}^{(0)}) + \boldsymbol{\Theta}_b(0, \delta\boldsymbol{S}_\tau, \bar{\boldsymbol{P}}^{(0)}) \\
+\boldsymbol{\Theta}_b(0, \boldsymbol{S}_\tau(\boldsymbol{p}^{(0)}), \delta\bar{\boldsymbol{P}}) + \boldsymbol{\Theta}_l(0, \bar{\boldsymbol{R}}) + \eta\boldsymbol{Y} \prec 0 \\
\boldsymbol{G}(\bar{\boldsymbol{P}}^{(0)} + \delta\bar{\boldsymbol{P}}) \prec \sigma^{(1)}\eta\boldsymbol{I} \\
\delta\boldsymbol{S}_\tau = \sum_{k=1}^{q} \frac{\partial \boldsymbol{S}_\tau}{\partial p_i}\Big|_{\boldsymbol{p}^{(0)}} \alpha_k \\
\boldsymbol{P}^{(0)} + \delta\boldsymbol{P}, \bar{\boldsymbol{P}}^{(0)} + \delta\bar{\boldsymbol{P}}, \boldsymbol{R}, \bar{\boldsymbol{R}} \succ 0, \eta > 0 \\
-\theta|\boldsymbol{p}^{(0)}| \le \boldsymbol{\alpha} \le \theta|\boldsymbol{p}^{(0)}| \\
\begin{bmatrix} 0.1\boldsymbol{P}^{(0)} & \delta\boldsymbol{P} \\ \delta\boldsymbol{P} & 0.1\boldsymbol{P}^{(0)} \end{bmatrix} \succ 0, \forall i = 1, \cdots, w \\
\begin{bmatrix} 0.1\bar{\boldsymbol{P}}^{(0)} & \delta\bar{\boldsymbol{P}} \\ \delta\bar{\boldsymbol{P}} & 0.1\bar{\boldsymbol{P}}^{(0)} \end{bmatrix} \succ 0, \forall i = 1, \cdots, w
\end{cases}
\tag{9-56}
$$

If it is feasible with solution $\boldsymbol{\alpha}, \delta\boldsymbol{P}, \delta\bar{\boldsymbol{P}}$, set $\boldsymbol{p}^{(1)} = \boldsymbol{p}^{(0)} + \boldsymbol{\alpha}$, $\boldsymbol{P}^{(1)} = \boldsymbol{P}^{(0)} + \delta\boldsymbol{P}$, $\bar{\boldsymbol{P}}^{(1)} = \bar{\boldsymbol{P}}^{(0)} + \delta\bar{\boldsymbol{P}}$ and go to step (2.3). Otherwise, set $\delta\sigma = \delta\sigma/2$ and repeat step (2.2).

(2.3) Solve the SDP (9-55) with $\sigma^{(0)}$ and $\boldsymbol{p}^{(0)}$ replaced with $\sigma^{(1)}$ and $\boldsymbol{p}^{(1)}$. If the problem is feasible, set $\sigma^{(0)} = \sigma^{(1)}$, $\boldsymbol{p}^{(0)} = \boldsymbol{p}^{(1)}$, $\boldsymbol{P}^{(0)} = \boldsymbol{P}^{(1)}$, and set $\delta\sigma = 2\delta\sigma$. Otherwise, and set $\theta = \theta/2$. Go to step (2.2).

Though phase 1 and phase 2 of the proposed algorithm have different design objectives, they strongly resemble each other in the underlying mechanism. Step (1.1) and (2.1) are the initialization steps to find a feasible $\beta^{(0)}$ or $\sigma^{(0)}$ along with the initial values of weighting matrices. Step (1.2) and (2.2) then linearize the problem around the current values of the decision variables by means of perturbations $\boldsymbol{\alpha}$, $\delta\boldsymbol{P}$, $\delta\bar{\boldsymbol{P}}$. To ensure the accuracy of the linearization, the perturbations should be kept reasonably small. Hence LMI constraints are introduced to express spectral norm limits $\|\delta\boldsymbol{P}\| \le 0.1\|\boldsymbol{P}\|$ and $\|\delta\bar{\boldsymbol{P}}\| \le 0.1\|\bar{\boldsymbol{P}}\|$. The smallness of control parameter change is ensured by $-\theta|\boldsymbol{p}^{(0)}| \le \boldsymbol{\alpha} \le \theta|\boldsymbol{p}^{(0)}|$ with a parameter θ adjustable during the iterative process. After obtaining the possibly favorable perturbations, the feasibility of the perturbed control parameters is verified and new values of weighting matrices are computed in step (1.3) and (2.3). It is suggested in [209] that $\boldsymbol{P}^{(0)}$ with smallest condition number works well in practice, therefore step (1.1)(2.1)(1.3)(2.3) minimize the ρ to suppress the condition number of the weighting matrices. The introduction of decision variable η significantly reduces numerical problems in solving the SDPs. In both phase 1 and phase 2, we implement a mechanism to dynamically adjust the expected improvement $\delta\beta$ and $\delta\sigma$ in each iteration, which makes the algorithm fully automatic and adaptive to different cases. Although we have employed the stability condition considering model reduction errors in presenting the algo-

rithm, the designers should also feel free to use the stability condition theorem 9.1 which does not have reduction error consideration only to bear in mind that this does not provide rigorous stability guarantee for the original full-order system.

9.5 Case Studies

In this section, we present case studies on several benchmark system given by the IEEE PES report [211]. Since the test systems are all well-known and all the information can be obtained from the full report [212], we omit the detailed description and diagrams and only mention the modification we have made to the systems. Our implementation of all the test systems are based on the Simulink model given by Abhinav K. Singh and Bikash C. Pal obtained from the website of Power System Dynamic Performance Committee. In addition, to fairly evaluate the LMI-based approach detailed in this chapter, the eigenanalysis based on Chebyshev discretization [44] is employed as a benchmark approach. The Chebyshev discretization method can approximate the eigenvalues of the time-delay systems at a convergence rate of $\mathcal{O}(N^{-N})$ [213], therefore it gives highly accurate eigenvalues when the number of discretization N is reasonably large, e.g. $N = 20$. In the following, the eigenvalues, damping factors, and delay margins calculated by Chebyshev discretization method are considered as the exact results. In our tests, YALMIP [118] is used as the modeling tool and Mosek [214] as the SDP solver.

9.5.1 Analysis

Stability analysis is the foundation of any meaningful design. Thus, before demonstrating the design power of the proposed method, we first present some stability analysis results on the 4-machine 2-area system. The configuration of controllers is as follows. Two PSSs are located at G1 and G3, and one WADC is installed to provide an additional input signal for the exitor of G1 using a remote signal of active power deviation of line 7-8. The remote signal is subject to a time delay. All the PSSs and WADCs are of the typical structure shown in Fig. 9-3. We first

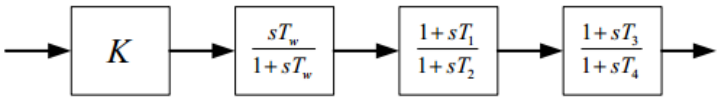


Figure 9-3 Transfer Function of PSSs and WADCs.

set the parameters of the controllers as those obtained from Phase 1 of our algorithm shown in the middle of Table 9-4. In our tests, we calculate the damping factor based delay margins, e.g. ϵ-delay margin is the maximal delay allowed for the system to have a damping factor at least ϵ. Table 9-2 shows the 10^{-4}-delay margins when varying the WADC gains. Column 2$\sim$4 contains the results obtained from the reduction error concerned stability condition (theorem 9.5) and column 5$\sim$6 include the results using the stability condition (theorem 9.1) without re-

Table 9-2 10^{-4}-Delay Margins (s) Calculated by Different Methods under Different WADC Gains on 2-area System

$-K_{wadc}$	ord	10	12	14	10	8	exact
	γ_0	$4e-2$	$7e-3$	$1e-3$	ignore error	ignore error	
	0.05	0.5409	0.5792	0.5838	0.5914	0.5891	0.5890
	0.1	0.4587	0.4902	0.4961	0.4983	0.5987	0.4969
	0.15	0.4203	0.4474	0.4535	0.4579	0.4584	0.4563
	0.2	0.3943	0.4220	0.4276	0.4310	0.4306	0.4285
	0.25	0.3730	0.3998	0.4051	0.4079	0.4081	0.4060
	0.3	0.3546	0.3806	0.3857	0.3874	0.3887	0.3866
	0.35	0.3379	0.3633	0.3683	0.3706	0.3712	0.3691
	0.4	0.3225	0.3473	0.3522	0.3528	0.3550	0.3530
	0.45	0.3079	0.3324	0.3371	0.3367	0.3398	0.3379
	0.5	0.2938	0.3182	0.3227	0.3210	0.3252	0.3235

duction error consideration. The final column of the Table lists the exact results obtained from Chebyshev discretization method. Table 9-3 further shows the 0.05-delay margins in the same way. It can be seen from both Tables that all the delay margin results are reasonably close to the exact values, which means that the ϵ-exponential stability criterion in theorem 9.1 contains very little conservatism and it is therefore of practical value to be used in control design. In fact, the ϵ-exponential stability criterion in this chapter is theoretically less conservative than those in [54, 55] also applied to calculate damping factor based delay margins due to the use of more accurate integral inequalities. In addition, all the delay margins obtained from the model reduction error concerned stability condition are always lower than the corresponding exact results. This means the model reduction error concerned stability condition, though using only the reduced system models, provides ϵ-exponential stability guarantee for the full-order systems. As the model orders of the reduced models increase, the reduction errors decrease and the delay margins from model reduction error concerned stability condition safely approach the exact results. In stark contrast, the delay margins obtained from the stability condition without reduction error consideration often exceed the exact values and those exaggerated results (highlighted by red color) exposes the intrinsic inability of stability analysis based on the reduced model to provide stability guarantee for the full-order system without considering the model reduction errors. Admittedly, the delay margins without reduction error consideration, though often overestimate, still stay very close to the exact values, while the delay margins considering model reduction errors are subjected to some degree of conservatism especially lower-order models are adopted.

Table 9-3 0.05-Delay Margins (s) Calculated by Different Methods under Different WADC Gains on 2-area System

	ord	10	12	14	10	8	exact
	γ_0	$4e-2$	$7e-3$	$1e-3$	ignore error	ignore error	
$-K_{wadc}$	0.05	0.4284	0.4592	0.4578	0.4725	0.4712	0.4702
	0.1	0.4042	0.4341	0.4402	0.4427	0.4430	0.4410
	0.15	0.3848	0.4130	0.4189	0.4222	0.4219	0.4197
	0.2	0.3677	0.3949	0.4004	0.4031	0.4034	0.4012
	0.25	0.3520	0.3782	0.3837	0.3859	0.3864	0.3843
	0.3	0.3372	0.3625	0.3680	0.3696	0.3705	0.3684
	0.35	0.3231	0.3478	0.3527	0.3542	0.3556	0.3535
	0.4	0.3096	0.3338	0.3384	0.3383	0.3412	0.3393
	0.45	0.2963	0.3202	0.3248	0.3225	0.3275	0.3255
	0.5	0.2834	0.3071	0.3115	0.3093	0.3141	0.3123

9.5.2 Design

The path-following method has been applied to design controllers on the 2-area system, the 3MIB system and the Brazilian 7-Bus system. On each test case, we obtain the initial parameters of the controllers either by the default values in the report [212] or by using the classical residue method for the system without time-delays [49]. Then we run the Phase 1 and Phase 2 of the path-following method to obtain the optimized control parameters.

The 2-area system has two local modes with frequency around 1.15 Hz, one within area 1 (G1 against G2) and another within area 2 (G3 against G4). There is an inter-area mode at the frequency of 0.61 Hz indicating area 1 swings against area 2. Without damping controllers, the inter-area mode is unstable and the local modes are lightly damped. The damping control configuration we implemented has been introduced in the last subsection, i.e. one PSS at G1, one PSS at G3 and one WADC at G1 using active power deviation of line 7-8 as the feedback signal. This remote signal is assumed to be subject to an 80ms time delay. All the controllers have the same structure as shown in Fig. 9-3. Hence We need to systematically design all the 6×3 control parameters in this case. Starting from the initial parameters, we run the proposed path-following algorithm with a 10-order reduced model targeting at improving the damping factor up beyond 0.1 and minimizing the H_2 dynamical performance. Fig. 9-4 show the progress of the iterations of Phase 1 and Phase 2 of the algorithm. Using 38 iterations, Phase 1 improves the damping factor from 0.0243 to 0.1018. We allow Phase 2 of the algorithm to run 100 iterations and observe that the H_2 metric has been reduced from 2.3958 to 0.5078. The initial parameters as well as the parameters after Phase 1 and Phase 2 fo the algorithm have been given in Table 9-4. The eigenanalysis using Chebyshev discretization demonstrates the signif-

icant improvement made by the algorithm shown in Fig. 9-7. It is shown the damping of the three electromechanical modes has been improved to a surprising extent by just coordinated tuning of the parameters of the three controllers. All the modes have damping ratios much larger than 10%. The two local modes achieve damping ratios of 27% and 30%, respectively. The damping ratio of the inter-area mode reaches 70%.

Similar studies are conducted on the 3MIB system and Brazilian 7-bus system as well. There are three critical modes for the 3MIB system. They characterize the oscillation of G1 against G2, G1&G2 against G3, and G1&G2&G3 against to infinite bus, respectively. Without damping control, the third mode is unstable and the second one is lightly damped. To improve damping, we install PSSs at G1 and G3, and one WADC is employed to provide an additional input signal for exitor of G1 using the active power deviation of line 4-5 as the feedback signal which has 40ms delay. The path-following algorithm runs with a 12-order reduced model to optimize the parameters of all the three controllers. The iteration process is shown in Fig. 9-5. The desired damping factor of 0.1 is achieved with 53 iterations of Phase 1. Phase 2 then is able to reduce the H_2 performance metric from 2.0295 to 0.7752. The eigenvalue plot in Fig. 9-8 illustrates the significant improvement of the damping of all critical modes. It shows all the modes have damping ratios great than 10% after optimization. The Brazilian 7-bus system is reported to be difficult to provide enough damping due to the existence of unstable transfer function zeros [215]. Any attempt to stabilize the system using single PSS is bound to fail. Report [212] provides a solution which installs four PSSs at G1$\sim$G4. This strategy improves the damping ratios beyond 5%. Here we propose another solution using only one PSS at G4 plus a WADC at G3 using the active power deviation of line 6-5 as the feedback signal. In our test, this remote signal is subjected to a time delay of 80ms. In the path-following algorithm, we use a 16-order reduced model and set target damping factor at 0.15. Fig. 9-6 shows the iteration process and eigenvalue plots of the system with obtained controllers are given in Fig. 9-9. After executing both Phases of the proposed algorithm, we are able to promote the damping factor up beyond 0.15 and achieve $>$10% damping ratios for all modes.

9.6 Conclusion

This chapter concerns the problem of time-domain design of damping controllers for time-delay power systems with a reduced-order system model. Instead of simply ignore the model reduction errors, we transform the full-order system as feedback interconnected of a reduced-order system and an error system, both involving time delays. We propose a model reduction method and a condition on the reduced system to ensure the ϵ-exponential stability of the full-order system. The controller design problem is then formulated as a H_2 performance minimization problem subject to nonlinear matrix inequality constraints. A path-following method is developed to solve the problem to local optimality. Case studies have demonstrated the

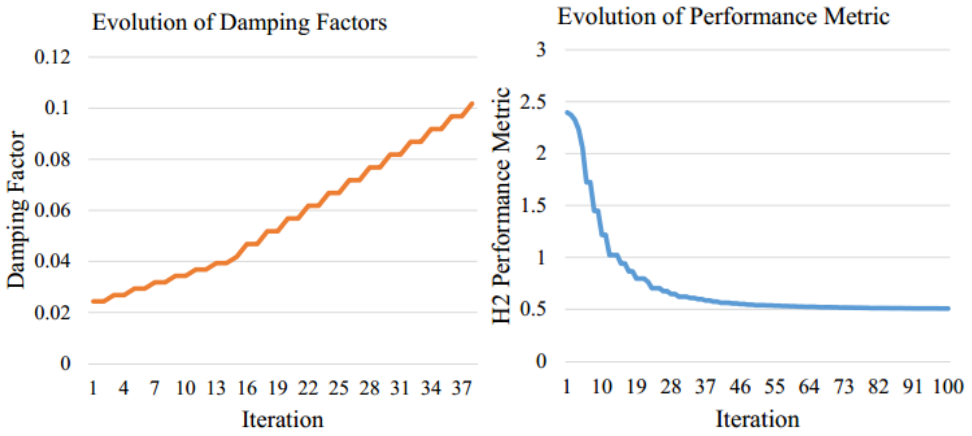


Figure 9-4 Evolution of damping factor ϵ and H_2 performance σ in the Phase 1 and Phase 2 of the path-following algorithm on the 2-area system.

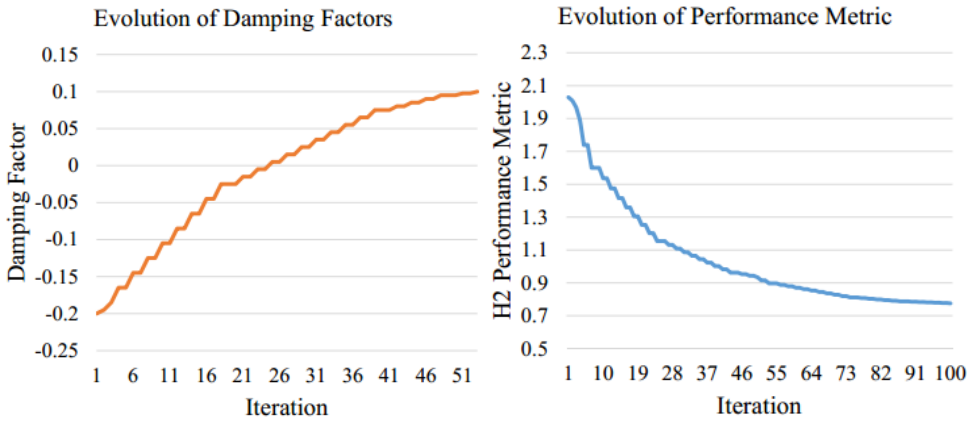


Figure 9-5 Evolution of damping factor ϵ and H_2 performance σ in the Phase 1 and Phase 2 of the path-following algorithm on the 3MIB system.

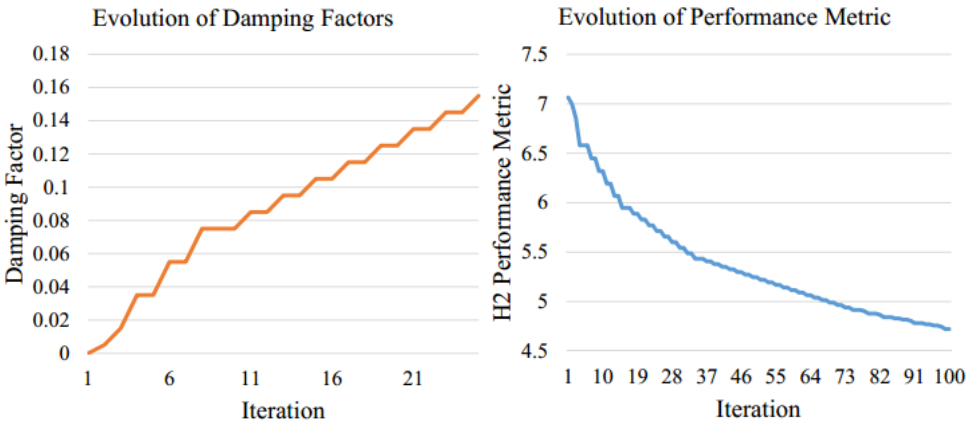
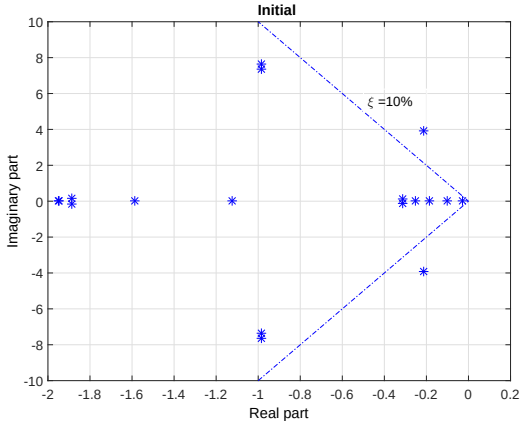
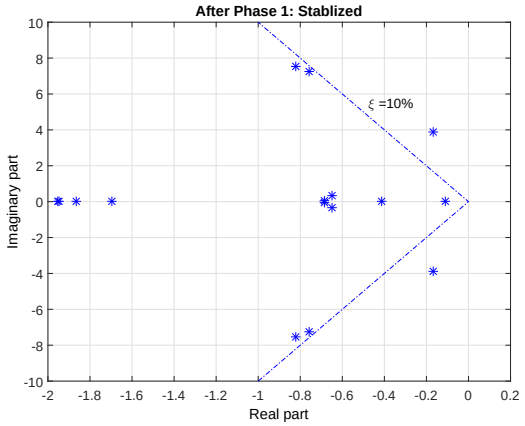


Figure 9-6 Evolution of damping factor ϵ and H_2 performance σ in the Phase 1 and Phase 2 of the path-following algorithm on the Brazilian 7-Bus system.

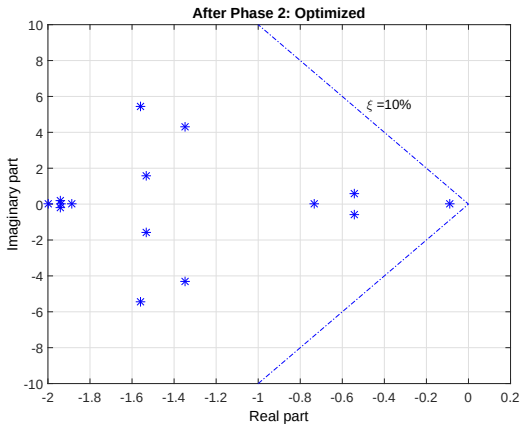
design power of the proposed method.



(a) Initial

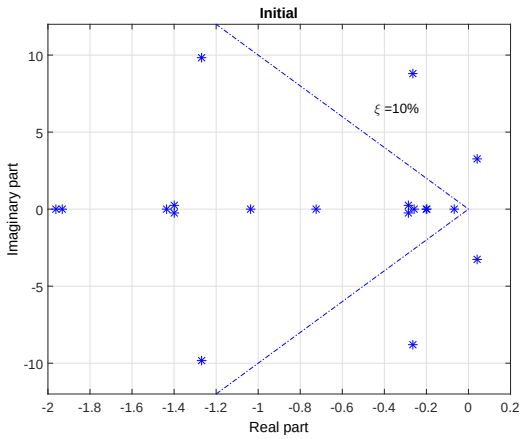


(b) After Phase 1

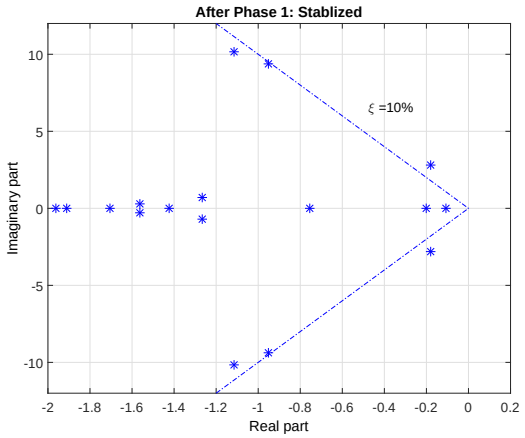


(c) After Phase 2

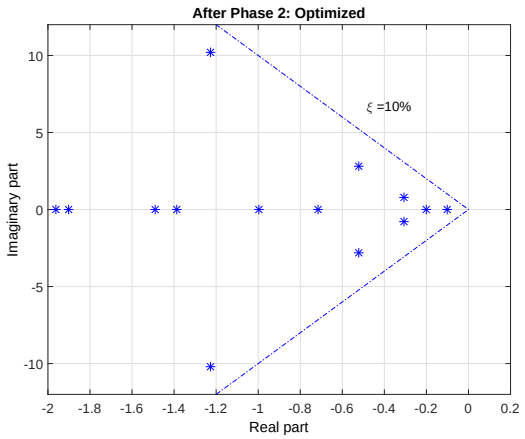
Figure 9-7 The eigenvalue plot of the test systems by chebyshev discretization method with the intial, stablized (Phase 1), and optimized (phase 2) control parameters of 2-area system.



(a) Initial

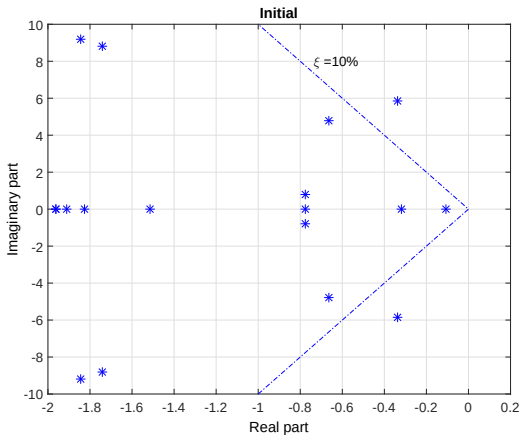


(b) After Phase 1

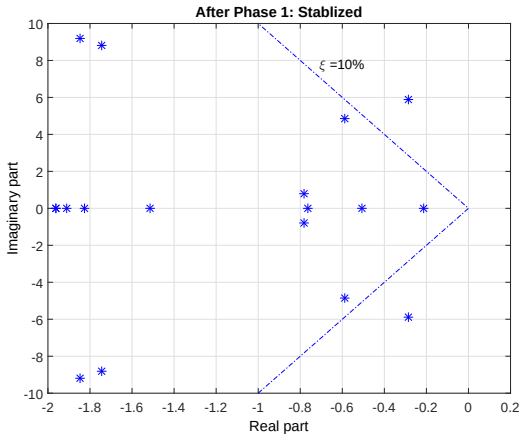


(c) After Phase 2

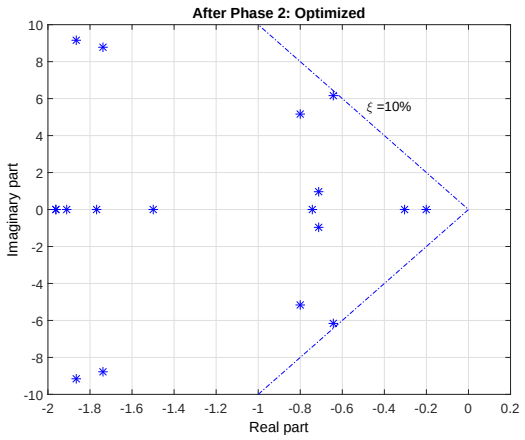
Figure 9-8 The eigenvalue plot of the test systems by chebyshev discretization method with the intial, stablized (Phase 1), and optimized (phase 2) control parameters of 3MIB system.



(a) Initial



(b) After Phase 1



(c) After Phase 2

Figure 9-9 The eigenvalue plot of the test systems by chebyshev discretization method with the intial, stablized (Phase 1), and optimized (phase 2) control parameters of Brazilian system.

Table 9-4 Controller Parameters of 2-area System

		K	Tw	T1	T2	T3	T4
Initial	wadc	-0.1	5	0.1016	0.6715	0.1016	0.6715
	pss G1	10	10	0.05	0.02	3	5.4
	pss G3	10	10	0.05	0.02	3	5.4
Phase 1	wadc	-0.0450	1.9348	0.0855	0.4173	0.0347	0.2479
	pss G1	4.2912	2.5446	0.0500	0.0106	0.5817	1.2518
	pss G3	7.0786	2.2558	0.0665	0.0236	0.5335	1.3324
Phase 2	wadc	-0.5078	1.7081	0.0051	1.2646	0.0027	0.0417
	pss G1	30.8739	1.3520	0.0519	0.0114	0.9134	0.4440
	pss G3	23.3095	0.6568	0.2958	0.0304	0.1328	0.1267

Table 9-5 Controller Parameters of 3MIB System

		K	Tw	T1	T2	T3	T4
Initial	wadc	-0.02	5	0.1129	1.0443	0.1129	1.0443
	pss G1	10	5	0.0572	0.238	0.0572	0.238
	pss G3	10	5	0.0572	0.238	0.0572	0.238
Phase 1	wadc	-0.0138	0.5977	0.0338	0.2890	0.0252	0.3210
	pss G1	7.1082	0.5833	0.0587	0.0834	0.0411	0.0921
	pss G3	5.7983	0.5883	0.0663	0.0678	0.0813	0.0886
Phase 2	wadc	-0.0620	1.2969	0.0004	0.4276	0.0002	0.2573
	pss G1	8.2428	1.0272	0.0575	0.0516	0.2253	0.4454
	pss G3	11.0236	0.2629	0.0654	0.0307	0.0647	0.0320

Table 9-6 Controller Parameters of 7-Bus System

		K	Tw	T1	T2	T3	T4
Initial	wadc G3	0.01	10	0.1320	0.2091	0.1320	0.2091
	pss G4	10	3	0.5200	0.0650	0.5200	0.0650
Phase 1	wadc G3	0.0089	4.9249	0.1241	0.2175	0.1147	0.2006
	pss G4	10.0527	1.9625	0.4566	0.0573	0.5075	0.0625
Phase 2	wadc G3	0.0186	5.6439	0.0010	0.1521	0.1253	0.1738
	pss G4	14.5633	3.0143	0.4765	0.0847	0.3279	0.0617

10 Conclusions and Prospects

10.1 Conclusions

In face of the challenges brought by the large-scale integration of renewable energy and increasingly close interconnection of power grids, this thesis has addressed some aspects of power system optimization and control under uncertainty and latency. Several key ideas are throughout the thesis. The first one is to use the distributionally robust method to deal uncertainties in power system operation. The second idea is to fully exploit the problem structures, e.g. sparsity and symmetry, to improve computational efficiency. We believe that problems appear in power systems are not generic ones, but present some special structures. If these structures are properly exploited, more efficient solution methods can be obtained. The third idea is to keep a balance between the convex optimization approach and the local search method. Recent publications in power system community have put too much emphasis on convex optimization models and algorithms so as to downplay the importance and effectiveness of local search methods. In fact, from a practitioner's viewpoint, a feasible local solution is much more favorable than an infeasible solution even if it is close to the global optimum, i.e. the feasibility is prior to the optimality. The main conclusions of this thesis are listed as follows.

(1) This thesis has applied the idea of distributionally robust optimization to unit commitment and co-optimization of energy-reserve-storage under uncertainties. Based on the nonparametric inference theory, a novel ambiguity set that contains the true probability distribution of uncertainties is constructed from observed historical data. The proposed model considers the worst-case distribution in the ambiguity set thereby achieves operational robustness. Moreover, the proposed ambiguity set shrinks to the true distribution as the number of historical data increases. Therefore, the conservatism of the solution can be reduced by incorporating more data. In addition, the scale of the optimization problem remains unchanged when using more data. Numerical studies demonstrate the favorable features of the proposed methods.

(2) This thesis has proposed a chance-constrained approximate AC-OPF under uncertainties based on distributionally robust optimization with Wasserstein metric. In order to overcome the flaws of the DC power flow model extensively used in stochastic OPF formulations, a tractable AC power flow model is developed by integrating the full AC power flow and a linear power flow models. Numerical studies have demonstrated the proposed OPF formulation has improved precision as well as good numerical tractability. The ambiguity set based on Wasserstein metric extracts reliable probabilistic information from historical data and the formulation immunizes the operation strategy against all distributions in the ambiguity set. Special problem structures are properly exploited in the reformulation of the distributionally

robust optimization problem to improve the scalability and efficiency of the numerical solution approach.

(3) This thesis has proposed a novel formulation and algorithm for FACTS devices allocation problems. Based on the sparse characteristics of device placement, FACTS allocation problems have been formulated as a sparsity-constrained OPF problem. An ADMM-IPM-STO algorithm, which combines the state-of-art algorithms in both sparse optimization and OPF, has been proposed to simultaneously determine the numbers, locations, setting values and types of FACTS devices. The algorithm is adaptive to various types of FACTS allocation problems and possesses the flexibility to use different sparsity-inducing norms to achieve desirable sparse features.

(4) The thesis has applied the moment-SOS approach to interval power flow analysis and the full AC multi-period OPF both of which can be formulated as non-convex polynomial optimization problems. To improve numerical tractability and efficiency, the chordal sparsity of the problem formulation is exploited to decompose the SDPs into smaller problems. In addition, ADMM has been employed to design a parallel algorithm for multi-period OPF problems by exploiting the separability of the problem formulation.

(5) The chordal sparsity and symmetry of the graph related to time-delay LFC loops have been exploited to improve the numerical tractability of DDSA. The graph-theoretic analysis provides guidance for restricting the structure of weighting matrices in DDSA, such that the LMIs possess chordal sparsity. The symmetry of LFC loops has been utilized to reduce the number of decision variables. At the price of introducing minor conservatism, case studies show the numerical tractability and computational efficiency have been improved by orders of magnitude.

(6) This paper has addressed the problem of time-domain design of damping controllers for time-delay power systems with a reduced-order system model. Instead of simply ignore the model reduction errors, we transform the full-order system as feedback interconnected of a reduced-order system and an error system, both involving time delays. We propose a model reduction method and a condition on the reduced system to ensure the ϵ-exponential stability of the full-order system. The controller design problem is then formulated as a H_2 performance minimization problem subject to nonlinear matrix inequality constraints. A path-following method is developed to solve the problem to local optimality.

10.2 Prospects

This thesis raises more questions than it answers. Some of the most important remaining problems are listed below:

(1) How can we deal with high dimension random variables in data-driven distributionally

robust optimization? When constructing an ambiguity set from the historical data, no matter what structure of ambiguity set is employed, the number of data points needed to achieve certain accuracy grows exponentially with the dimension of the random variables. This is the phenomenon we call the statistical curse of dimensionality. Does this mean that it is inherently impossible for us to learn reliable information about the high-dimensional uncertainties from finite data? If so, what's the reasonable expectation when our optimization problems involve high-dimensional uncertainties?

(2) It is observed that the second-order moment relaxation obtains the exact global solutions of the OPF and IPF problems in many cases we have tested. What's the underlying mechanism of this phenomenon? What physical or topological properties of the power grids have led to the good performance of second-order moment relaxation? Of the many semidefinite constraints in the second-order moment relaxation, which constraints play the key role in tightening the convex relaxation, and can other constraints be neglected without affecting the exactness of the solutions?

(3) In the delay-dependent stability analysis of time-delay power systems, even if the sparsity is exploited, the current solvers still cannot solve real-world large-scale problems involving hundreds of generators. We know that the frequency domain eigenanalysis approach is able to deal with very large systems due to the development of sparse linear algebra. Therefore, can we establish a connection between the time-domain LMI-based approach and the frequency-domain methods and hence reformulate the LMIs as equivalent eigenvalue problems to accelerate their solution?

(4) In Chapter 9 of this thesis, we have proposed a new concept call dissipative model reduction which not only cares about the input-output approximation of reduced systems but only enforce exponential dissipativity requirement on the error systems. Can we extend this idea to enforce other properties on the error systems to help reduce conservatism of the stability criterion considering model reduction errors?

(5) Convergence of the path-following method is in doubt.

(6) In many decomposition algorithms to solve power system optimization problems including the ones presented in Chapter 5 and 7 of this thesis, a nonlinear programming subproblem needs to be solved iteratively usually using IPM. In such algorithms, any two consecutive iterates are very close to each other. Therefore, huge computational savings can be gained by warm-starting the IPM in each iteration using the last iteration point. However, the IPM is notoriously difficult to warm-start. How to reliably warm-start IPM is still an open question.

Appendix A Adaptive Barrier Filter Line-search IPM for Optimal Power Flow with FACTS Devices

This appendix Chapter gives a detailed account of the interior point method employed in Chapter 5 to solve the optimal power flow subproblem (5-7) with FACTS devices.

A.1 Introduction

Modern power systems demand stronger self-control ability to meet various technical and economic requirements of market participants. Thus large quantities of FACTS devices have been installed to facilitate steady and dynamic control of power systems in recent two decades [150] . Consequently conventional algorithms to obtain system control strategies are challenged by those developments of power systems.

Optimal power flow (OPF) problems have been proposed for half a century and the research on the formulations and algorithms of OPF has been experiencing continuous development. Because various new elements, such as FACTS devices, are continuously added into power systems, up to now OPF has developed into a special research field with plentiful contents [216].

Interior point method (IPM) is one of the most successful algorithms applied to OPF problems among various methods. Specially, it has almost become a standard method to solve OPF problems in recent years. The primal-dual IPM (PD IPM) [217], along with its high-order variants the predictor corrector IPM (PC IPM) [138] and the multiple centrality corrections IPM (MCC IPM) [137] are the most widely applied and extensively discussed algorithms to OPF problems. They have successfully solved conventional OPF problems on not only standard test systems but also large-scale real-world systems [218]. However, when applied to OPF problems with FACTS devices, the reliability of all the above three IPMs should be seriously questioned. In our numerical experience, above IPMs sometimes get stuck at some non-optimal points with the step-lengths becoming extremely small and finally fail to achieve a local optimum especially when there are a considerable number of FACTS devices in the system. The installation of FACTS devices not only increases the variable dimension but also intensifies the nonlinearity of OPF problems. Theoretically and practically, no IPM can guarantee convergence in general nonlinear OPF problems.

In order to overcome or at least alleviate this drawback of existing OPF algorithms, a new OPF algorithm with the latest knowledge of nonlinear optimization theory is put into practice in this chapter. Dealing with OPF problem with FACTS devices formulated in [219], this chapter introduces three new techniques to improve the robustness of IPM. The adaptive barrier

parameter update strategy [220] reasonably controls the decrease of the barrier parameter and prevents iterates prematurely approaching feasibility boundary. The filter line-search method [221] efficiently avoids unfavorable long steps and ensures new iterate to progress toward the solution meanwhile. The feasibility restore phase [222] is taken as a remedy to restore the algorithm in case convergence difficulty occurs.

Case studies on hundreds of randomly generated OPF problems with FACTS devices show that the novel algorithm is more robust than previous ones, and gives faster performance compared with MCC IPM. About ten thousands of numerical tests on both standard test systems and large-scale real-world systems have demonstrated all three strategies above largely increase the possibility of the algorithm to successfully solve the problems.

A.2 General OPF Formulation With FACTS Devices

In this chapter, the OPF problem is formulated in rectangular coordinates with current mismatch equations. Generators and loads are modelled as complex current injections at their buses. All FACTS devices are modelled as parametric complex current injections at related buses [219]. The motivation for these choices is to facilitate the calculation of the second order derivatives. First, in the most general case, there are up to three series controllable parameters associated with each line, which leads to a very high order power mismatch equation using polar coordinates, and solving the second order derivation in such formulation is very difficult. Second, with parametric current injections used to depict the effects of FACTS devices, the nodal admittance matrix stays constant during the optimization process.

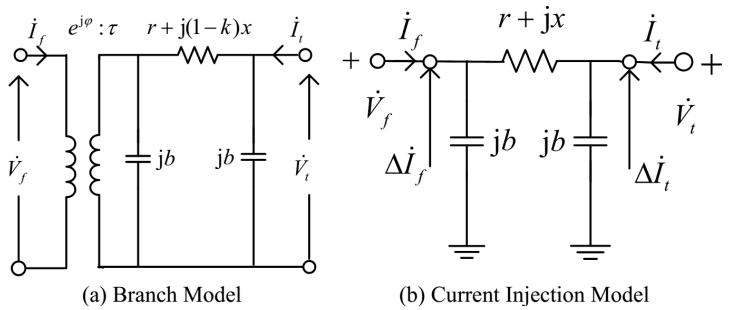
A.2.1 Branch Model

A general branch model is shown in Fig.A-1 (a) which is similar to that in MATPOWER [**?**]. r, x and b are transmission line parameters. $\dot{V}_f$,$\dot{V}_t$, $\dot{I}_f$ and $\dot{I}_t$ are complex voltages and currents at “from” and “to” ends of the branch. The series controllable parameters$k$, φ and τ are used to describe the effects of TCSC, TCPS and ULTC, respectively. Shunt compensation devices can be modelled as extra susceptance at their buses so that they are not shown in Fig. A-1.

It is expected that the effects of all the controllable parameters are represented by the auxiliary current injections at the two ends shown in Fig. A-1 (b). Combining Fig. A-1(a) with Fig. A-1 (b), the auxiliary parametric complex current injections are obtained as

$$\begin{bmatrix} \Delta\dot{I}_f \\ \Delta\dot{I}_t \end{bmatrix} = \begin{bmatrix} y-\tau^2\tilde{y}+\mathrm{j}b(1-\tau^2) & -y+\tau e^{\mathrm{j}\varphi}\tilde{y} \\ -y+\tau e^{-\mathrm{j}\varphi}\tilde{y} & y-\tilde{y} \end{bmatrix} \begin{bmatrix} \dot{V}_f \\ \dot{V}_t \end{bmatrix} \tag{A-1}$$

where $\tilde{y}=1/[r+\mathrm{j}(1-k)x]$. Equation (1) exhibits that $\Delta\dot{I}_f$ and $\Delta\dot{I}_t$ are the functions of series



(a) Branch Model (b) Current Injection Model

Figure A-1 Branch Modelling

controllable parameters and complex voltages at both ends of the line.

A.2.2 OPF Formulation

Current mismatch equations are chosen as equality constraints. In this formulation, bus voltages and generator current injections are taken as state variables. For the ith bus, the current mismatch equation is given by

$$\dot{I}_{Gi} + (\dot{S}_{Li}/\dot{V}_i)^* + \Delta\dot{I}_i - \sum_{k=1}^{N} Y_{ik}\dot{V}_k = 0 \tag{A-2}$$

where $\dot{I}_{Gi}$ and $\dot{S}_{Li}$ are the complex current injection of generators and the complex power injection of loads at the ith bus, respectively, Y_{ik} is the (i, k)th element in the constant nodal admittance matrix of the original network shown in Fig. A-1 (b), $\Delta\dot{I}_i$ is the sum of the complex current injections induced by all the controllable devices related to the ith bus.

$$\Delta\dot{I}_i = \sum_{f\in i}\Delta\dot{I}_f + \sum_{t\in i}\Delta\dot{I}_t - \mathrm{j}B_{shi}\dot{V}_i \tag{A-3}$$

The last term in (3) denotes the complex current injection induced by shunt compensation devices at the ith bus. The following steady state security constraints of power systems are considered in our study:

$$P_{G_i\min} \le P_{G_i} \le P_{G_i\max},\ 1 \le i \le n_G \tag{A-4}$$

$$Q_{G_i\min} \le Q_{G_i} \le Q_{G_i\max},\ 1 \le i \le n_G \tag{A-5}$$

$$V_{i\min}^2 \le \left|\dot{V}_i\right|^2 \le V_{i\max}^2,\ 1 \le i \le n_b \tag{A-6}$$

$$0 \le \left|\dot{I}_i\right|^2 \le I_{i\max}^2,\ 1 \le i \le n_l \tag{A-7}$$

$$\tau_{i\min} \le \tau_i \le \tau_{i\max},\ 1 \le i \le n_l \tag{A-8}$$

$$B_{sh_i\min} \le B_{sh_i} \le B_{sh_i\max},\ 1 \le i \le n_b \tag{A-9}$$

$$\varphi_{i\min} \leq \varphi_i \leq \varphi_{i\max},\ 1 \leq i \leq n_l \tag{A-10}$$

$$k_{i\min} \leq k_i \leq k_{i\max},\ 1 \leq i \leq n_l \tag{A-11}$$

The frequently-used objectives in OPF are minimizing generation cost, maximizing loadability and minimizing transmission losses, etc. In the general case, the complete decision vector consists of the loadability factor, real and imaginary parts of bus voltages, real and imaginary parts of generator current injections, UTLC ratios, and setting values of all the FACTS devices. To a specific problem, the decision vector $\boldsymbol{x}$ may be chosen as a sub-vector of the complete decision vector by decision-makers. To sum up, the OPF formulation with FACTS devices can be written in a compact form as (5)

$$\begin{aligned} &\min_{\boldsymbol{x}}\ f(\boldsymbol{x}) \\ &\text{s.t. } \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \\ &\boldsymbol{g}_{\min} \leq \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{g}_{\max} \end{aligned} \tag{A-12}$$

where decision vector $\boldsymbol{x}$ is n_x dimensional, $\boldsymbol{h}(\boldsymbol{x})$ is an n_h dimensional function specified by (2), $\boldsymbol{g}(\boldsymbol{x})$ is an n_g dimensional function given by (4) and $f(\boldsymbol{x})$ is an objective function.

A.3 Adaptive Barrier Filter Line-search IPM

To develop a more robust IPM for OPF problems with FACTS devices, several critical improvements are made to existing OPF algorithms. First, the barrier parameter is not decreased until the current μ-barrier problem is solved to certain accuracy. Second, the centering parameter is chosen by an adaptive parameter update strategy in which the centering parameter is determined based on the minimization of a clear-cut quality function instead of simple heuristics used in previous algorithms. Third, corrector steps used in PC and MCC algorithms are abandoned. Forth, the filter line-search method is adopted to generate the next iterate. Fifth, when the line-search procedure cannot make sufficient progress and the step length becomes too small, a feasibility restore phase is initiated to obtain a new acceptable iterate. In this section, we first sketch out the framework of the new IPM and then go into the details of the adaptive barrier parameter update strategy, the filter line-search method and the feasibility restore phase.

A.3.1 Outline of the Interior Point Method

IPMs replace the original problem (5) with a sequence of μ-barrier problems:

$$\begin{aligned}
&\min_{\boldsymbol{z}} \ \varphi_\mu(\boldsymbol{z}) = f(\boldsymbol{x}) - \mu\sum_{i=1}^{r}\ln(l_i) - \mu\sum_{i=1}^{r}\ln(u_i) \\
&\text{s.t. } \boldsymbol{h}(\boldsymbol{x}) = 0 \\
&\boldsymbol{g}_{\min} - \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{l} = 0,\ \boldsymbol{l} > 0 \\
&\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{g}_{\max} + \boldsymbol{u} = 0,\ \boldsymbol{u} > 0
\end{aligned} \tag{A-13}$$

where $\boldsymbol{l}$and$\boldsymbol{u}$are slack variables, and$\mu$is the positive barrier parameter. The augmented Lagrange function associated with (6) is constructed as

$$\begin{aligned}
&L_\mu(\boldsymbol{z},\boldsymbol{y}) = f(\boldsymbol{x}) - \mu\sum_{i=1}^{r}\ln(l_i) - \mu\sum_{i=1}^{r}\ln(u_i) + \boldsymbol{\lambda}^{\mathrm{T}}\boldsymbol{h}(\boldsymbol{x}) \\
&+ \boldsymbol{v}^{\mathrm{T}}\left(\boldsymbol{g}_{\min} - \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{l}\right) + \boldsymbol{w}^{\mathrm{T}}\left(\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{g}_{\max} + \boldsymbol{u}\right)
\end{aligned} \tag{A-14}$$

where$\boldsymbol{\lambda}$, bmv and $\boldsymbol{w}$ are Lagrange multipliers for equality and inequality constraints. Denote$\boldsymbol{y} = [\boldsymbol{\lambda}^{\mathrm{T}}, \boldsymbol{v}^{\mathrm{T}}, \boldsymbol{w}^{\mathrm{T}}]^{\mathrm{T}}$.$\boldsymbol{z}and\boldsymbol{y}$are n_z dimensional primal and n_y dimensional dual variables, respectively. The KKT optimality condition for μ-barrier problem (6) takes the form

$$\begin{bmatrix} \nabla_{\boldsymbol{x}} L_\mu \\ \nabla_{\boldsymbol{l}} L_\mu \\ \nabla_{\boldsymbol{u}} L_\mu \\ \nabla_{\boldsymbol{\lambda}} L_\mu \\ \nabla_{\boldsymbol{v}} L_\mu \\ \nabla_{\boldsymbol{w}} L_\mu \end{bmatrix} = \begin{bmatrix} \nabla f(\boldsymbol{x}) + \nabla\boldsymbol{h}(\boldsymbol{x})\boldsymbol{\lambda} - \nabla\boldsymbol{g}(\boldsymbol{x})\boldsymbol{v} + \nabla\boldsymbol{g}(\boldsymbol{x})\boldsymbol{w} \\ [\boldsymbol{l}]\boldsymbol{v} - \mu\boldsymbol{e} \\ [\boldsymbol{u}]\boldsymbol{w} - \mu\mathrm{e} \\ \boldsymbol{h}(\boldsymbol{x}) \\ \boldsymbol{g}_{\min} - \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{l} \\ \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{g}_{\max} + \boldsymbol{u} \end{bmatrix} = 0 \tag{A-15}$$

where $\mathrm{e} = [1, \cdots, 1]^{\mathrm{T}}$, $[\boldsymbol{l}]$ denotes the diagonal matrix with the elements of the vector $\boldsymbol{l}$ on the diagonal. Applying the Newton method to nonlinear algebraic equation (8), the Newton direction from current iterate$(\boldsymbol{z}^k, \boldsymbol{y}^k)$ can be obtained by solving the following linear system

$$\nabla^2 L_\mu(\boldsymbol{z}^k, \boldsymbol{y}^k) \begin{bmatrix} \Delta\boldsymbol{z}^k \\ \Delta\boldsymbol{y}^k \end{bmatrix} = -\nabla L_\mu(\boldsymbol{z}^k, \boldsymbol{y}^k) \tag{A-16}$$

Then the next iterate is given by

$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k + \beta\alpha_{\boldsymbol{z}}^{\max}\Delta\boldsymbol{z}^k \tag{A-17}$$

$$\boldsymbol{y}^{k+1} = \boldsymbol{y}^k + \alpha_{\boldsymbol{y}}^{\max}\Delta\boldsymbol{y}^k \tag{A-18}$$

The maximum step lengths $\alpha_{\boldsymbol{z}}^{\max}$ and $\alpha_y^{\max}$ are obtained by using the fraction-to-the-boundary rule to guarantee the positivity of $\boldsymbol{l}$, $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$.

$$\alpha_{\boldsymbol{z}}^{\max} = \min\left\{\tau\min_i\left\{\left.\frac{-l_i^k}{\Delta l_i^k}\right| \Delta l_i^k < 0,\ \left.\frac{-u_i^k}{\Delta u_i^k}\right| \Delta u_i^k < 0\right\}, 1\right\} \tag{A-19}$$

$$\alpha_{\boldsymbol{y}}^{\max}=\min\left\{\tau\min_i\left\{\left.\frac{-v_i^k}{\Delta v_i^k}\right|\Delta v_i^k<0,\ \ \left.\frac{-w_i^k}{\Delta w_i^k}\right|\Delta w_i^k<0\right\},1\right\} \tag{A-20}$$

where $\tau\in(0,1)$ is a safety factor determined by $\tau=\max\{0.99,1-\mu\}$ in our implementation. Note that $\alpha_{\boldsymbol{y}}^{\max}$ and $\beta\alpha_{\boldsymbol{z}}^{\max}$ are the actual step sizes for dual and primal variables, respectively. In conventional algorithms, $\beta$is always set to 1. In this chapter, it is determined by a backtracking filter line-search procedure which will be discussed later on. Instead of changing barrier parameter at each iterate, the value of μ is kept intact until an approximate solution to the μ-barrier problem is obtained. The optimality error of a μ-barrier problem is defined as

$$E_\mu(\boldsymbol{z}^k,\boldsymbol{y}^k)=\max\{\varepsilon_1,\ \varepsilon_2,\ \varepsilon_3\} \tag{A-21}$$

where

$$\varepsilon_1=\max\{\|\boldsymbol{h}(\boldsymbol{x})\|_\infty,\|\boldsymbol{g}_{\min}-\boldsymbol{g}(\boldsymbol{x})+\boldsymbol{l}\|_\infty,\|\boldsymbol{g}(\boldsymbol{x})-\boldsymbol{g}_{\max}+\boldsymbol{u}\|_\infty\} \tag{A-22}$$

$$\varepsilon_2=\frac{\|\nabla f(\boldsymbol{x})+\nabla\boldsymbol{h}(\boldsymbol{x})\boldsymbol{\lambda}-\nabla\boldsymbol{g}(\boldsymbol{x})\boldsymbol{v}+\nabla\boldsymbol{g}(\boldsymbol{x})\boldsymbol{w}\|_\infty}{1+\|\boldsymbol{x}\|_2+\|\boldsymbol{\lambda}\|_2+\|\boldsymbol{v}\|_2+\|\boldsymbol{w}\|_2} \tag{A-23}$$

$$\varepsilon_3=\frac{\|[\boldsymbol{v}]\boldsymbol{l}-\mu\mathsf{e}\|_1+\|[\boldsymbol{w}]\boldsymbol{u}-\mu\mathsf{e}\|_1}{1+\|\boldsymbol{x}\|_2} \tag{A-24}$$

ε_1,$\varepsilon_2$and$\varepsilon_3$represent the feasibility, optimality and centrality of the current iterate, respectively. For a given value of μ, in this chapter, the μ-barrier problem (6) is solved to satisfy

$$E_\mu(\boldsymbol{z}^*,\boldsymbol{y}^*)\le k_\mu\mu \tag{A-25}$$

where$k_\mu$is a positive parameter determined by decision -makers. Then a new barrier parameter is obtained from the adaptive barrier parameter update strategy. The overall algorithm terminates if the approximate solution $(\boldsymbol{z}^*,\boldsymbol{y}^*)$ satisfies

$$E_0(\boldsymbol{z}^*,\boldsymbol{y}^*)\le\varepsilon_{\text{tol}} \tag{A-26}$$

where tol is an error tolerance.

The main steps of the novel IPM, named adaptive barrier filter-line search (ABFLS) IPM in this chapter, are shown in Fig. A-2, in which the details will be discussed next.

A.3.2 Adaptive Barrier Parameter Update Strategy

The barrier parameter is usually chosen to be proportional to the current complementarity gap

$$\mu=\frac{\sigma}{2n_g}\left(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{l}+\boldsymbol{w}^{\mathrm{T}}\boldsymbol{u}\right) \tag{A-27}$$

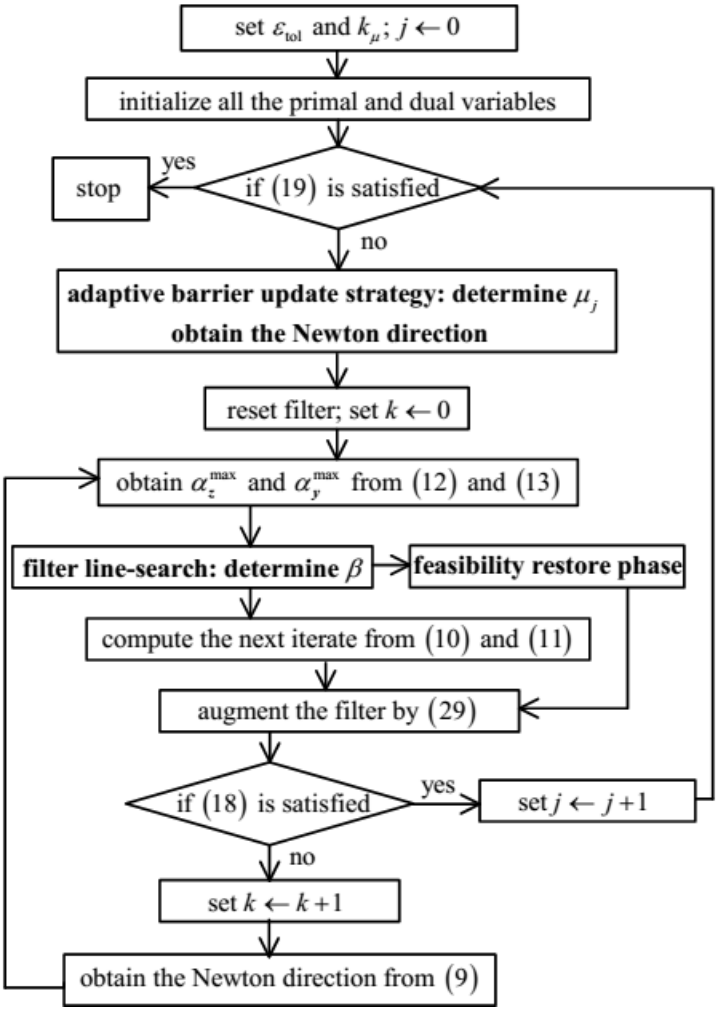


Figure A-2 Flow Chart of ABFLS IPM

where$\sigma$is the centering parameter. σhas significant influence on the convergence of IPM algorithms. If σ is selected, the barrier parameterμ, the KKT equation (8), and further the Newton direction (9) are successively determined. In the PD algorithm,$\sigma$is simply set to a constant. In the PC and MCC algorithms, some simple heuristics are employed to updateσ. The adaptive barrier parameter update strategy is originally proposed in [220], and it is firstly adopted in OPF problems with FACTS devices in this chapter. Note that when algorithms converge, the norm of mismatch vector of equation (9) $\|\nabla L_0(\boldsymbol{z},\boldsymbol{y})\|_2^2$approaches 0. Therefore, the selected σ should provide minimum value of quality function

$$q_N(\sigma) = \left\|\nabla L_0(\boldsymbol{z}^{k+1}(\sigma), \boldsymbol{y}^{k+1}(\sigma))\right\|_2^2 \tag{A-28}$$

where $\boldsymbol{z}^{k+1}(\sigma)$ and $\boldsymbol{y}^{k+1}(\sigma)$ can be calculated by successively applying (20), (9), (12), (13), (10) and (11). The value of βin (10) is set to 1 at this stage. However, the evaluation of $q_N(\sigma)$ is too expensive because it needs to compute the new iterate for every trial value ofσ. To circumvent this obstacle, (21) can be expressed as (22) considering$f(\boldsymbol{x})$,$\boldsymbol{h}(\boldsymbol{x})and\boldsymbol{g}(\boldsymbol{x})$ are all

linear functions, since the solution of the linear equation can be exactly achieved by just one Newton step.

$$\begin{aligned} q_L(\sigma) &= \left[1-\alpha_{\boldsymbol{y}}^{\max}(\sigma)\right]^2 \left\|\nabla_{\boldsymbol{x}} L_\mu(\boldsymbol{z}^k,\boldsymbol{y}^k)\right\|_2^2 \\ &+ \left[1-\alpha_{\boldsymbol{z}}^{\max}(\sigma)\right]^2 \left\|\nabla_{\boldsymbol{y}} L_\mu(\boldsymbol{z}^k,\boldsymbol{y}^k)\right\|_2^2 \\ &+ \left\|[\boldsymbol{l}^{k+1}(\sigma)]\boldsymbol{v}^{k+1}(\sigma)\right\|_2^2 + \left\|[\boldsymbol{u}^{k+1}(\sigma)]\boldsymbol{w}^{k+1}(\sigma)\right\|_2^2 \end{aligned} \tag{A-29}$$

Although the quality function$q_N(\sigma)$is substituted by $q_L(\sigma)$in purpose of reducing computation burden, the trend induced by$q_L(\sigma)$is the same as by$q_N(\sigma)$. Practically, a larger value ofσ, i.e. a conservativeμupdate, will lead to relatively longer step lengths and smaller values of the first two terms in (22). However a large complementary gap will be produced meanwhile, so that the values of the last two terms in (22) are larger. The optimal value of σcan balance long step lengths and smaller complementary gap. To search the optimal σ, new iterate with respect to trial σmust be computed. Note that the Hessen matrix is not related to $\sigma$and the RHS of the Newton equation (9) is a linear function of σ, therefore

$$\Delta\boldsymbol{z}^k(\sigma) = \Delta\boldsymbol{z}^k(0) + \sigma(\Delta\boldsymbol{z}^k(1) - \Delta\boldsymbol{z}^k(0)) \tag{A-30}$$

$$\Delta\boldsymbol{y}^k(\sigma) = \Delta\boldsymbol{y}^k(0) + \sigma(\Delta\boldsymbol{y}^k(1) - \Delta\boldsymbol{y}^k(0)) \tag{A-31}$$

After obtaining $\Delta\boldsymbol{z}^k(0)$,$\Delta\boldsymbol{y}^k(0)$,$\Delta\boldsymbol{z}^k(1)$ and $\Delta\boldsymbol{y}^k(1)$,$\Delta\boldsymbol{z}^k(\sigma)$ and $\Delta\boldsymbol{y}^k(\sigma)$ can be easily computed for any trial value of σ from (23) and (24), and it is only twice that the linear system (9) is solved for $\sigma = 0$and 1. Also, it only takes a few vector operations to evaluate the step length $\alpha_{\boldsymbol{z}}^{\max}(\sigma)$ and $\alpha_{\boldsymbol{y}}^{\max}(\sigma)$ according to (12) and (13). Thus, $q_L(\sigma)$ can be cheaply computed and applied to a one-dimensional search scheme.

A.3.3 Filter Line-search Method

Taking the maximum step size $\alpha_{\boldsymbol{z}}^{\max}$ is not always beneficial to the convergence of the algorithm especially when the current iterate is far away from the optimal point. It is a common strategy to determine the actual step size by applying a backtracking line-search procedure to explore a decreasing sequence of the step size coefficient β with an initial value of 1. In the context of solving the μ-barrier problem in IPM, the filter method, originally proposed by Fletcher and Leyffer [223], is implemented in [8, 9]. The basic idea of the filter method is to find a new iterate providing progress in terms of either objective function $\varphi_\mu(\boldsymbol{z})$ or constraints violation

$$\theta(\boldsymbol{z}) = \left\| \begin{matrix} \boldsymbol{h}(\boldsymbol{x}) \\ \boldsymbol{g}_{\min} - \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{l} \\ \boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{g}_{\max} + \boldsymbol{u} \end{matrix} \right\|_\infty \tag{A-32}$$

compared with current and previous iterates. Therefore $(\varphi_\mu(\boldsymbol{z}^p),\ \theta(\boldsymbol{z}^p))$ pair defines a "taboo region":

$$\mathrm{R}^p = \{(\varphi_\mu, \theta) |\, \varphi_\mu \geq (1-\gamma_\varphi)\varphi_\mu(\boldsymbol{z}^p) \text{ and } \theta \geq (1-\gamma_\theta)\theta(\boldsymbol{z}^p)\} \tag{A-33}$$

where γ_φ and γ_θ are small safety constant parameters and $\gamma_\varphi = \gamma_\theta = 10^{-5}$ in our implementation. The algorithm maintains a "filter", which is the union of taboo region corresponding to the current and previous iterates

$$\mathrm{F}^k = \bigcup_{p=1}^{k} \mathrm{R}^p \tag{A-34}$$

The trial point $\boldsymbol{z}^{k+1}(\beta)$ is accepted as the next iterate if

$$\left(\varphi_\mu(\boldsymbol{z}^{k+1}(\beta)),\ \theta(\boldsymbol{z}^{k+1}(\beta))\right) \notin \mathrm{F}^k \tag{A-35}$$

and then the filter is augmented to form the taboo region of the next iterate

$$\mathrm{F}^{k+1} = \mathrm{F}^k \cup R^{p+1} \tag{A-36}$$

Otherwise, $\boldsymbol{z}^{k+1}(\beta)$ is rejected and $\beta \leftarrow \beta/2$. Then the new trial point is calculated and tested. This process is repeated until an acceptable β is found or $\beta\alpha_{\boldsymbol{z}}^{\max}$ is less than $\alpha_{\boldsymbol{z}}^{\min}$. If the latter happens, the algorithm gives up the line-search procedure and initiates the feasibility restore phase to compute a new acceptable iterate. If the current μ-barrier problem is successfully solved to expected accuracy, the filter is reset to an empty set.

A.3.4 Feasibility Restore Phase

The feasibility restore phase [222] aims at finding a new iterate acceptable to the current filter by reducing the constraints violation. It is intuitively formulated as an optimization problem to find a feasible point being the closest to the current point:

$$\begin{aligned} &\min_{\boldsymbol{z}} \left\| \boldsymbol{D}(\boldsymbol{z} - \boldsymbol{z}^k) \right\|_2^2 \\ &\text{s.t. } \boldsymbol{h}(\boldsymbol{x}) = 0 \\ &\boldsymbol{g}_{\min} - \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{l} = 0,\ \boldsymbol{l} > 0 \\ &\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{g}_{\max} + \boldsymbol{u} = 0,\ \boldsymbol{u} > 0 \end{aligned} \tag{A-37}$$

where $\boldsymbol{D}$ is a diagonal scaling matrix whose elements are

$$D_i = min\left\{1, 1/|\boldsymbol{z}_i^k|\right\},\ i = 1, 2, \cdots, n_z \tag{A-38}$$

Since the objective is a positive definite quadratic function, the optimal solution to problem (30) is usually a strict local minimum in the manifold defined by the equality constraints of (30), which makes problem (30) relatively easier to solve. The augmented Lagrangian Method [153] with projected Newton steps [224] is applied to solve this problem in this chapter. The

Table A-1 Problem Types

Problems	FACTS Devices Installed	Problems	FACTS Devices Installed
P1	10 SVC	P6	10 TCPS
P2	20 SVC	P7	10 SVC & 10 TCSC
P3	5 TCSC	P8	10 SVC & 10 TCPS
P4	10 TCSC	P9	10 TCSC & 10 TCPS
P5	5 TCPS	P10	10 SVC & 10 TCSC & 10 TCPS

algorithm terminates and reverts to regular IPM iterations once an acceptable point is obtained.

A.4 Case Studies

Comparative studies among PD, PC, MCC and the proposed ABFLS IPMs are reported in this section. Comparison is made from both efficiency and robustness perspectives on not only IEEE standard test systems but also real-world systems. Some difficult operational conditions are also considered to assess the robustness of the proposed method. In addition, to the proposed ABFLS IPM, the relationship between problem scale and its performance is discussed. In all the tests, we take maximizing the loadability as the objective function to assess the methods under possibly heavier load level. In such condition, the number of binding inequality constraints should be much higher than that under lower load level, which results in more difficult OPF problems. In this way, we can differentiate all the studied methods in robustness. All the four algorithms are coded in MATLAB running on a personal computer with Intel Core i5 1.80-GHz CPU and 3.85 GB of RAM. All system data is extracted from MATPOWER. The centering parameter for PD algorithm is set to 0.2. In PC and MCC algorithms, centering parameters are determined by the same heuristic used in [137]. The MCC algorithm runs with the same parameter as in [225]: $\beta_{\min} = 0.1$, $\beta_{\max} = 10$, $\delta_{\alpha} = 0.2$, $\varepsilon_{\alpha} = 0.03$ and $K_{\max} = 5$. In the ABFLS algorithm, $k_{\mu} = 50$. The convergence tolerance ε_{tol} is set to 10-4. In the adaptive barrier update strategy, the golden section method is applied in the interval of $\sigma \in [0, 10]$ and terminates after 12 evaluations of the quality function, or the search interval becomes smaller than 10-2.

A.4.1 Case studies on IEEE Standard Test Systems

We first evaluate all the algorithms by applying them to maximizing loadability problems on IEEE test systems with different types and numbers of FACTS devices. Ten types of problems are defined and shown in Table A-1 and all tests are carried out on these problems.

In the first numerical experiment, P1 to P10 type problems are generated on every test system by randomly installing corresponding FACTS devices. With all the four algorithms applied to those problems, the numbers of iterations to convergence are recorded. The results

Table A-2 Number of Iterations on IEEE-57 System

	Problems									
Algorithms	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
PD	58	33	79	F	37	37	43	F	49	48
PC	23	33	24	21	26	30	64	32	F	126
MCC	12	11	10	F	9	10	13	12	84	176
ABFLS	12	11	18	19	13	14	17	14	21	32

Table A-3 Number of Iterations on IEEE-118 System

	Problems									
Algorithms	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
PD	93	161	45	54	46	41	37	52	42	360
PC	29	27	49	67	30	44	33	39	66	53
MCC	67	31	51	83	73	39	37	61	26	96
ABFLS	22	19	19	19	23	22	26	26	30	63

of IEEE-57, -118 and -300 bus systems are shown in Table A-2, A-3, and A-4, respectively. The case that convergence is not achieved within 500 iterations is considered a failure and denoted as “F”.

Previous papers [5, 14] report that the MCC algorithm generally outperforms the PC algorithm and PC algorithm compares favorably with the PD algorithm in efficiency. This phenomenon has also been observed in our calculation especially on IEEE-57 system (Table A-2) which is the smallest test system in our experiment. The performance of the PC and MCC algorithms on P1-P3 and P5-P8 problems on the IEEE-57 system shows their acceleration effect compared with the PD algorithm. However, as the increase of system scale and the number of FACTS devices, the acceleration effect of the PC and MCC algorithms becomes rather uncertain. The experiments on the IEEE-118 and IEEE -300 systems show that they do not necessarily give better performance than the PD algorithm. Even on the IEEE-57 system, their unfavorable performance on P9 and P10 reveals that the placement of several FACTS devices (especially a considerable number of TCSCs) may deteriorate the convergence process of the

Table A-4 Number of Iterations on IEEE-300 System

	Problems									
Algorithms	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
PD	52	49	F	116	48	F	F	F	F	129
PC	F	127	F	F	F	F	F	77	F	F
MCC	111	52	144	42	39	33	F	23	F	F
ABFLS	28	25	38	36	45	24	28	22	52	47

Table A-5 Success Rate Comparison on IEEE-300 System

	Problems					
Algorithm	P1	P2	P3	P4	P5	P6
PD	65%	67%	67%	64%	62%	65%
PC	40%	49%	38%	29%	37%	33%
MCC	91%	94%	89%	75%	93%	94%
ABFLS	100%	100%	98%	92%	100%	100%

PC and MCC algorithms. In our perspective, the fundamental reason for this phenomenon is that the PC and MCC algorithms for nonlinear OPF problems are direct extension of Mehrotra's [226] and Gondzio's [227] methods originally proposed for linear programming. Hence there is no sound theoretical foundation to guarantee their performance in nonlinear problems. In addition, the introduction of a considerable number of FACTS devices may intensify the nonlinearity of the problem formulation and thus impairs the acceleration effect of the PC and MCC algorithms. The performance of the proposed ABFLS algorithm is very stable. It needs slightly more iterations than MCC algorithm on simple problems. However, when the problems become harder, the ABFLS algorithm outperforms the MCC algorithm in terms of the number of iterations. The first numerical experiment shows that IPMs may fail to converge on some hard problems, which motivates the need to study the robustness of all the four algorithms. In the second experiment, we randomly generate 100 sets of problems (P1-P6) on IEEE-300 system. Applying all the four algorithms to these problems, we record the percentage of problems successfully solved within 500 iterations by each algorithm. The results of this experiment are provided in Table A-5.

Table A-5 shows that PC is the most unreliable one among the four algorithms. This observation agrees with the numerical results and discussion in [220] which reveals that some inconsistency of corrector steps may significantly increase the complementary gap and lead to convergence failure. Table A-5 also exhibits the relatively high robustness of the ABFLS algorithm among all the four algorithms. This robust performance stems from the adaptive barrier parameter update strategy, the filter line-search method and the feasibility restore phase. No algorithm can guarantee 100Note that Table A-5 reports the MCC algorithm is much more reliable than the PD and PC algorithms but only a bit less reliable than ABFLS algorithm. Comparing the MCC and ABFLS algorithms in terms of efficiency, we list the average iteration numbers and CPU time of the two algorithms if certain problem is successfully solved in Table A-6. This shows that ABFLS outperforms MCC in terms of both robustness and efficiency.

Table A-6 Average Iteration Number and CPU Time on IEEE-300 System

Algorithm		Problems					
		P1	P2	P3	P4	P5	P6
MCC	iter	80.8	69.7	80.5	75.1	68.4	77.0
	time	8.3s	7.2s	8.6s	8.0s	7.3s	8.2s
ABFLS	iter	40.8	40.0	51.5	69.4	42.5	46.5
	time	5.0s	4.9s	6.9s	9.3s	5.7s	6.2s

A.4.2 Case studies on Real-world Systems

The ABFLS IPM is also applied to large-scale real-world systems, including a 2736-bus, a 3012-bus and a 3120-bus systems. In our experiment, P1 type problems are randomly generated in all the above three systems. PD, PC, MCC and the proposed ABFLS IPMs are all employed to solve each problem.

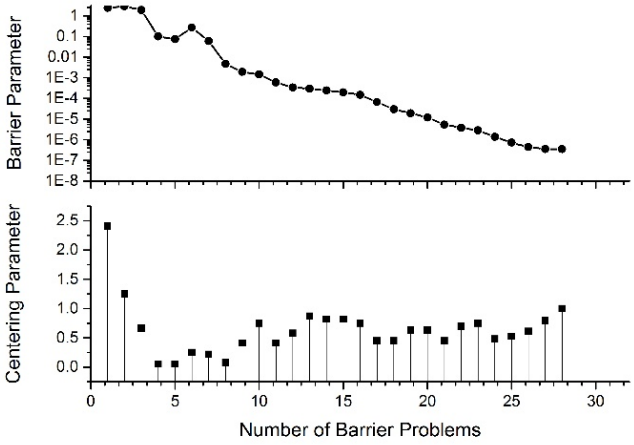


Figure A-3 Centering and Barrier Parameters

For saving space, only the performance of ABFLS IPM on the 2736-bus system is shown with details of the convergence process in Fig. A-3, Fig. A-4 and Fig. A-5. In this case, 28 μ-barrier problems are generated and solved to expected accuracy, and the number of iterations for the overall problem is 30. Fig. A-3 shows that the barrier parameter μ decreases non-monotonously since centering parameter σ is greater than 1 at some iterations. Those larger values of σ help to avoid small step lengths at certain iterations, which improves the robustness of the algorithm. Fig. A-4 presents the convergence of three criteria and the values of β at each iteration. Those less than 1 values of β prevent unfavorable long steps and benefit the convergence process. The evolution of objective function is shown in Fig. A-5. Since the initial point (base load flow) is infeasible (violating some voltage magnitude limits), the loadability first drop to a lower value to find a feasible point and then level up to reach the maximal value

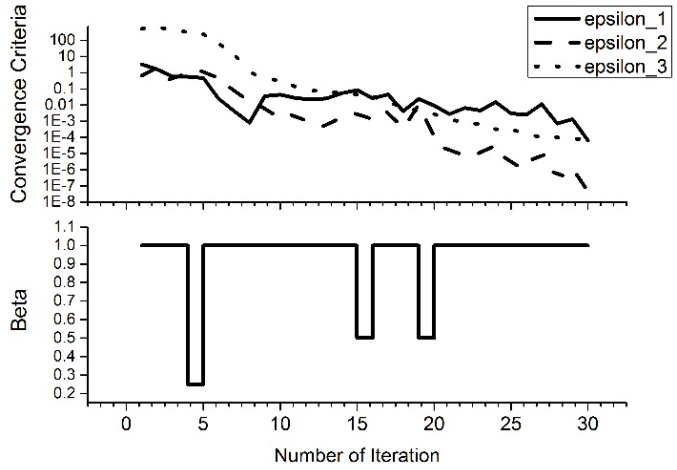


Figure A-4 Convergence Criteria and Values of Beta

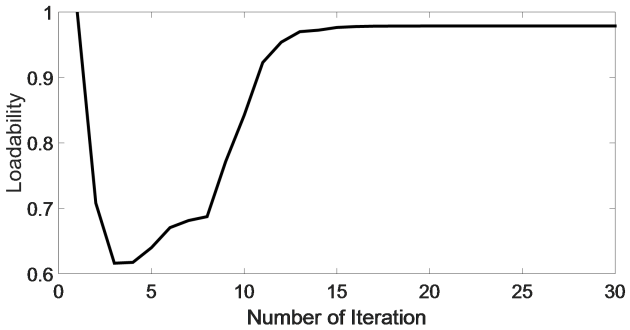


Figure A-5 Evolution of Objective Function (Loadability)

slightly smaller than the initial value. Table A-7 reports the comparison among PD, PC, MCC and ABFLS on real-world systems. It shows that ABFLS is generally more robust and needs less iterations.

A.4.3 Case Study under Difficult Operation Condition

The OPF problems can be made more difficult not only by the introduction of FACTS devices but also unfavourable operation conditions, especially very tight operational limits [218]. In this subsection, we exam the robustness of the proposed method compared with

Table A-7 Number of Iterations of Different IPMs on Real-world Systems

Algorithm	Problems		
	Case2736	Case3012	Case3120
PD	63	F	63
PC	55	69	90
MCC	65	F	44
ABFLS	30	42	38

other methods by gradually tightening the operational limits. Two groups of tests have been conducted. In the first group, we set the voltage magnitude limits of all the buses $[1-\delta, 1+\delta]$ p.u. with the parameter gradually decreasing from 0.05 to a value with which all these methods fail to converge. In the second group, we set the line current magnitude limits $[0, \tau I_{i\max}]$, $i = 1, \cdots, n_l$ with a common parameter gradually decreasing from 1 to a value with which all these methods fail to converge. The experiments are conducted on IEEE 39-bus system with no FACTS device. Comparison is made among PC, MCC and the proposed ABFLS algorithms. The results are reported in Fig. A-6 and Fig. A-7. In the first group of tests, all three methods fail to solve the problem with the parameter . The numbers of iteration needed by each method with parameter decreasing from 0.05 to 0.006 by 0.001 in each step are presented in Fig. A-6. This figure shows that ABFLS algorithm is able to solve all the problems while PC and MCC fail at some problems. Particularly, for $\delta = 0.007$, both PC and MCC fail whereas ABFLS successfully solve this extremely difficult problem to required accuracy although with 451 iterations. In the second groups of tests, the parameter decreases from 1 to 0.28 and all the three methods fail to solve the problem with $\tau = 0.28$. Fig. A-7 pictures the numbers of iteration needed for each method to converge for parameters from 1 to 0.3 by decreasing 0.02 in each step. Also, we can observe from Fig. A-7 that the ABFLS solves all the problems while PC and MCC fail at some cases. These results coincide with our previous observation that the proposed ABFLS algorithm needs slightly more iterations than MCC on simple problems but is generally more stable and robust on hard problems. In addition, PC is the least robust among the three.

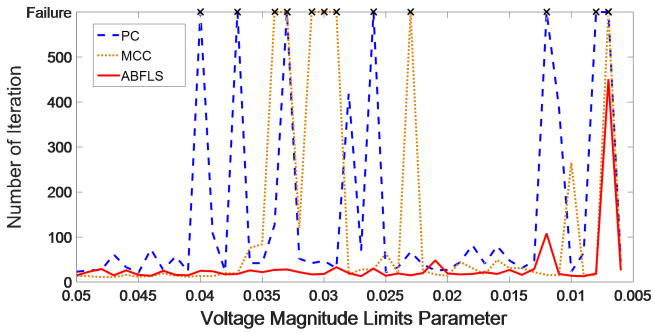


Figure A-6 Number of Iteration to Convergence with Different δ

A.4.4 Relationship between System Scale and Performance

We generate another set of P1 type problems on 14-, 30-, 39-, 57-, 118-, 300-, 2736-, 3012-, and 3120-bus systems. With the proposed ABFLS IPM applied to those problems, the numbers of iterations and time for all problems are presented in Table A-8.

The number of iterations is not sensitive to problem scale, while the time consumed per iteration is nearly propositional to the problem scale. Hence there is roughly a linear or at least

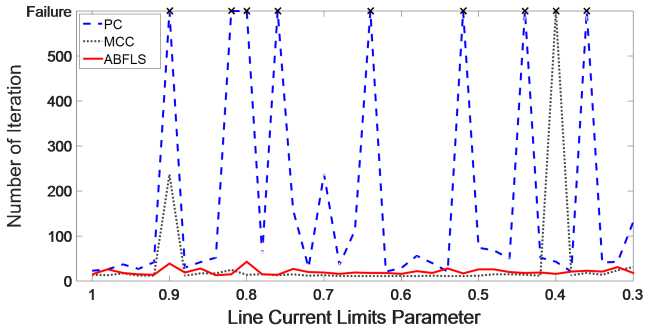


Figure A-7 Number of Iteration to Convergence with Different τ

Table A-8 Performance of the ABFLS IPM on Different Systems

Systems	14	30	39	57	118	300	2736	3012	3120
iter	16	18	29	15	44	39	37	36	50
time (s)	0.8	1.0	1.3	0.9	3.1	4.8	113.3	181.0	261.8

a super-linear relationship between the overall time and the problem scale.

A.5 Conclusion

In the engineering computation practice, the convergence reliability of exist IPM algorithms is not high enough when they are applied to solve the OPF problems, especially there are a large quantity of FACTS devices in power systems. In view of this, three measures, namely the adaptive barrier update strategy, the filter line-search method and the feasibility restore phase, have been simultaneously introduced in the conventional primal-dual interior point method framework to enhance the robustness of OPF algorithms in this paper. About ten thousands of numerical tests on both standard systems and large-scale real-world systems for this ABFLS IPM have demonstrated its convergence reliability. Comparative case studies show that the performance of the PC and MCC algorithms degrades as system scale and the FACTS device number increase. The proposed ABFLS algorithm is reliable and efficient, and outperforms the famous PD, PC and MCC algorithms in both robustness and efficiency.

References

[1] Renewables 2017: global status report[J]. Environmental Policy Collection. 2017.

[2] Koivisto M, Sørensen P, Maule P, et al. Needs for flexibility caused by the variability and uncertainty in wind and solar generation in 2020, 2030 and 2050 scenarios[M]. [S.l.]: DTU Wind Energy, 2017.

[3] 刘振亚. 中国电力与能源 [M]. 北京: 中国电力出版社, 2012.

[4] 刘振亚. 全球能源互联网 [M]. 北京: 中国电力出版社, 2015.

[5] Yu H, Chung C, Wong K, et al. Probabilistic load flow evaluation with hybrid latin hypercube sampling and cholesky decomposition[J]. IEEE Transactions on Power Systems. 2009, 24(2):661–667.

[6] Hajian M, Rosehart W D, Zareipour H. Probabilistic power flow by monte carlo simulation with latin supercube sampling[J]. IEEE Transactions on Power Systems. 2013, 28(2):1550–1559.

[7] Allan R, Da Silva A L, Burchett R. Evaluation methods and accuracy in probabilistic load flow solutions[J]. IEEE Transactions on Power Apparatus and Systems. 1981, PAS-100(5):2539–2546.

[8] Zhang P, Lee S T. Probabilistic load flow computation using the method of combined cumulants and gram-charlier expansion[J]. IEEE Transactions on Power Systems. 2004, 19(1):676–682.

[9] Hu Z, Wang X. A probabilistic load flow method considering branch outages[J]. IEEE Transactions on Power Systems. 2006, 21(2):507–514.

[10] Su C L. Probabilistic load-flow computation using point estimate method[J]. IEEE Transactions on Power Systems. 2005, 20(4):1843–1851.

[11] Morales J M, Perez-Ruiz J. Point estimate schemes to solve the probabilistic power flow[J]. IEEE Transactions on Power Systems. 2007, 22(4):1594–1601.

[12] Borkowska B. Probabilistic load flow[J]. IEEE Transactions on Power Apparatus and Systems. 1974, 93(3):752–759.

[13] Da Silva A, Arienti V, Allan R. Probabilistic load flow considering dependence between input nodal powers[J]. IEEE Transactions on Power Apparatus and Systems. 1984, 6(PAS-103):1524–1530.

[14] Fan M, Vittal V, Heydt G T, et al. Probabilistic power flow studies for transmission systems with photovoltaic generation using cumulants[J]. IEEE Transactions on Power Systems. 2012, 27(4): 2251–2261.

[15] Ren Z, Li W, Billinton R, et al. Probabilistic power flow analysis based on the stochastic response surface method[J]. IEEE Transactions on Power Systems. 2015, PP(99):1–9.

[16] Wang Z, Alvarado F L. Interval arithmetic in power flow analysis[J]. IEEE Transactions on Power Systems. 1992, 7(3):1341–1349.

[17] Dimitrovski A, Tomsovic K. Boundary load flow solutions[J]. IEEE Transactions on Power Systems. 2004, 19(1):348–355.

[18] Vaccaro A, Canizares C A, Villacci D. An affine arithmetic-based methodology for reliable power flow analysis in the presence of data uncertainty[J]. IEEE Transactions on Power Systems. 2010, 25 (2):624–632.

[19] Vaccaro A, Cañizares C A, Bhattacharya K. A range arithmetic-based optimization model for power flow analysis under interval uncertainty[J]. IEEE Transactions on Power Systems. 2013, 28(2):1179–1186.

[20] Ding T, Bo R, Li F, et al. Interval power flow analysis using linear relaxation and optimality-based bounds tightening (obbt) methods[J]. IEEE Transactions on Power Systems. 2015, 30(1):177–188.

[21] Zheng Q, Wang J, Pardalos P, et al. A decomposition approach to the two-stage stochastic unit commitment problem[J]. Annals of Operations Research. 2013, 210(1):387–410.

[22] Wang Q, Guan Y, Wang J. A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output[J]. IEEE Transactions on Power Systems. Feb 2012, 27(1):206–215.

[23] Takriti S, Birge J, Long E. A stochastic model for the unit commitment problem[J]. IEEE Transactions on Power Systems. Aug 1996, 11(3):1497–1508.

[24] Carpentier P, Gohen G, Culioli J C, et al. Stochastic optimization of unit commitment: a new decomposition framework[J]. IEEE Transactions on Power Systems. May 1996, 11(2):1067–1073.

[25] Wu L, Shahidehpour M, Li T. Stochastic security-constrained unit commitment[J]. IEEE Transactions on Power Systems. May 2007, 22(2):800–811.

[26] Pappala V, Erlich I, Rohrig K, et al. A stochastic model for the optimal operation of a wind-thermal power system[J]. IEEE Transactions on Power Systems. May 2009, 24(2):940–950.

[27] Zheng Q, Wang J, Liu A. Stochastic optimization for unit commitment: A review[J]. IEEE Transactions on Power Systems. July 2015, 30(4):1913–1924.

[28] Ozturk U, Mazumdar M, Norman B. A solution to the stochastic unit commitment problem using chance constrained programming[J]. IEEE Transactions on Power Systems. Aug 2004, 19(3):1589–1598.

[29] Jiang R, Wang J, Guan Y. Robust unit commitment with wind power and pumped storage hydro[J]. IEEE Transactions on Power Systems. May 2012, 27(2):800–810.

[30] Bertsimas D, Litvinov E, Sun X, et al. Adaptive robust optimization for the security constrained unit commitment problem[J]. IEEE Transactions on Power Systems. Feb 2013, 28(1):52–63.

[31] Jiang R, Wang J, Zhang M, et al. Two-stage minimax regret robust unit commitment[J]. IEEE Transactions on Power Systems. Aug 2013, 28(3):2271–2282.

[32] Lee C, Liu C, Mehrotra S, et al. Modeling transmission line constraints in two-stage robust unit commitment problem[J]. IEEE Transactions on Power Systems. May 2014, 29(3):1221–1231.

[33] An Y, Zeng B. Exploring the modeling capacity of two-stage robust optimization: Variants of robust unit commitment model[J]. IEEE Transactions on Power Systems. Jan 2015, 30(1):109–122.

[34] Zhao C, Guan Y. Unified stochastic and robust unit commitment[J]. IEEE Transactions on Power Systems. Aug 2013, 28(3):3353–3361.

[35] Zhao C, Guan Y. Data-driven stochastic unit commitment for integrating wind generation[J]. IEEE Transactions on Power Systems. 2016, 31(4):2587–2596.

[36] Xiong P, Jirutitijaroen P, Singh C. A distributionally robust optimization model for unit commitment considering uncertain wind power generation[J]. IEEE Transactions on Power Systems. 2016, PP (99):1–11.

[37] Wei W, Liu F, Mei S. Distributionally robust co-optimization of energy and reserve dispatch[J]. IEEE transactions on Sustainable Energy. 2016, 7(1):289–300.

[38] Wei W, Wang J, Mei S. Dispatchability maximization for co-optimized energy and reserve dispatch with explicit reliability guarantee[J]. IEEE Transactions on Power Systems. 2016, 31(4):3276–3288.

[39] Bian Q, Xin H, Wang Z, et al. Distributionally robust solution to the reserve scheduling problem with partial information of wind power[J]. IEEE Transactions on Power Systems. 2015, 30(5):2822–2823.

[40] Wang Z, Bian Q, Xin H, et al. A distributionally robust co-ordinated reserve scheduling model considering cvar-based wind power reserve requirements[J]. IEEE transactions on Sustainable Energy. 2016, 7(2):625–636.

[41] Lubin M, Dvorkin Y, Backhaus S. A robust approach to chance constrained optimal power flow with renewable generation[J]. IEEE Transactions on Power Systems. 2016, 31(5):3840–3849.

[42] Zhang Y, Shen S, Mathieu J. Distributionally robust chance-constrained optimal power flow with uncertain renewables and uncertain reserves provided by loads[J]. IEEE Transactions on Power Systems. 2016, PP(99):1–1.

[43] Bienstock D, Chertkov M, Harnett S. Chance-constrained optimal power flow: risk-aware network control under uncertainty[J]. SIAM Review. 2014, 56(3):461–495.

[44] Milano F, Anghel M. Impact of time delays on power system stability[J]. IEEE Transactions on Circuits & Systems I Regular Papers. 2012, 59(4):889–900.

[45] Milano F, Dassios I. Small-signal stability analysis for non-index 1 hessenberg form systems of delay differential-algebraic equations[J]. IEEE Transactions on Circuits & Systems I Regular Papers. 2016, 63(9):1521–1530.

[46] Ye H, Liu Y, Zhang P. Efficient eigen-analysis for large delayed cyber-physical power system using explicit infinitesimal generator discretization[J]. IEEE Transactions on Power Systems. 2016, 31(3): 2361–2370.

[47] Şahin Sönmez, Ayasun S, Nwankpa C O. An exact method for computing delay margin for stability of load frequency control systems with constant communication delays[J]. IEEE Transactions on Power Systems. 2015, 31(1):370–377.

[48] Pagola F L, Perez-Arriaga I J, Verghese G C. On sensitivities, residues and participations: applications to oscillatory stability analysis and control[J]. IEEE Transactions on Power Systems. 1989, 4(1):278–285.

[49] Aboul-Ela M E, Sallam A, McCalley J D, et al. Damping controller design for power system oscillations using global signals[J]. IEEE Transactions on Power Systems. 1996, 11(2):767–773.

[50] Yang N, Liu Q, Mccalley J D. Tcsc controller design for damping interarea oscillations[J]. IEEE Transactions on Power Systems. 2002, 13(4):1304–1310.

[51] Yao W, Jiang L, Wu Q, et al. Delay-dependent stability analysis of the power system with a wide-area damping controller embedded[J]. IEEE Transactions on Power Systems. 2011, 26(1):233–240.

[52] Yao W, Jiang L, Wen J, et al. Wide-area damping controller of facts devices for inter-area oscillations considering communication time delays[J]. IEEE Transactions on Power Systems. 2013, 29(1):318–329.

[53] Li J, Chen Z, Cai D, et al. Delay-dependent stability control for power system with multiple time-delays[J]. IEEE Transactions on Power Systems. 2016, 31(3):2316–2326.

[54] Yang B, Sun Y. Damping factor based delay margin for wide area signals in power system damping control[J]. IEEE Transactions on Power Systems. 2013, 28(3):3501–3502.

[55] Yang B, Sun Y. A novel approach to calculate damping factor based delay margin for wide area damping control[J]. IEEE Transactions on Power Systems. 2014, 29(6):3116–3117.

[56] Yao W, Jiang L, Wen J, et al. Wide-area damping controller of facts devices for inter-area oscillations considering communication time delays[J]. IEEE Transactions on Power Systems. 2014, 29(1):318–329.

[57] Jiang L, Yao W, Wu Q, et al. Delay-dependent stability for load frequency control with constant and time-varying delays[J]. IEEE Transactions on Power Systems. 2012, 27(2):932–941.

[58] Zhang C K, Jiang L, Wu Q, et al. Further results on delay-dependent stability of multi-area load frequency control[J]. IEEE Transactions on Power Systems. 2013, 28(4):4465–4474.

[59] Wu M, He Y, She J H. Stability analysis and robust control of time-delay systems[M]. [S.l.]: Springer, 2010.

[60] Yu X, Tomsovic K. Application of linear matrix inequalities for load frequency control with communication delays[J]. IEEE Transactions on Power Systems. 2004, 19(3):1508–1515.

[61] Bevrani H. Robust power system frequency control: volume 85[M]. [S.l.]: Springer, 2009.

[62] Zhang C K, Jiang L, Wu Q, et al. Delay-dependent robust load frequency control for time delay power systems[J]. IEEE Transactions on Power Systems. 2013, 28(3):2192–2201.

[63] Bian Q, Xin H, Wang Z, et al. Distributionally robust solution to the reserve scheduling problem with partial information of wind power[J]. IEEE Transactions on Power Systems. 2015, 30(5):2822–2823.

[64] Wei W, Liu F, Mei S. Distributionally robust co-optimization of energy and reserve dispatch[J]. IEEE transactions on Sustainable Energy. 2016, 7(1):289–300.

[65] Lavaei J, Low S. Zero duality gap in optimal power flow problem[J]. IEEE Transactions on Power Systems. Feb 2012, 27(1):92–107.

[66] Gan L, Li N, Topcu U, et al. Exact convex relaxation of optimal power flow in radial networks[J]. IEEE transactions on Automatic Control. Jan 2015, 60(1):72–87.

[67] Madani R, Sojoudi S, Lavaei J. Convex relaxation for optimal power flow problem: Mesh networks[J]. IEEE Transactions on Power Systems. Jan 2015, 30(1):199–211.

[68] Deng J, Li C, Zhang X P. Coordinated design of multiple robust facts damping controllers: A bmi-based sequential approach with multi-model systems[J]. IEEE Transactions on Power Systems. 2015, 30(6):3150–3159.

[69] Ban M, Yu J, Shahidehpour M, et al. Integration of power-to-hydrogen in day-ahead security-constrained unit commitment with high wind penetration[J]. Journal of Modern Power Systems and Clean Energy. 2017, 5(3):337.

[70] Wasserman L. All of nonparametric statistics[M]. [S.l.]: Springer Science & Business Media, 2006.

[71] Goldman M, Kaplan D M. Evenly sensitive KS-type inference on distributions[R]. [S.l.]: [s.n.], 2015.

[72] Devroye L, Wise G L. Detection of abnormal behavior via nonparametric estimation of the support[J]. SIAM Journal on Applied Mathematics. 1980, 38(3):480–488.

[73] Ba Í, Cuevas A, Justel A, et al. Set estimation and nonparametric detection[J]. Canadian Journal of Statistics. 2000, 28(4):765–782.

[74] Jabr R A. Adjustable robust opf with renewable energy sources[J]. IEEE Transactions on Power Systems. 2013, 28(4):4742–4751.

[75] Jabr R, Karaki S, Korbane J. Robust multi-period opf with storage and renewables[J]. IEEE Transactions on Power Systems. Sept 2015, 30(5):2790–2799.

[76] Lorca A, Sun X A, Litvinov E, et al. Multistage adaptive robust optimization for the unit commitment problem[J]. Operations Research. 2016, 64(1):32–51.

[77] Lorca A, Sun X. Multistage robust unit commitment with dynamic uncertainty sets and energy storage[J]. IEEE Transactions on Power Systems. 2016, PP(99):1–1.

[78] Wiesemann W, Kuhn D, Sim M. Distributionally robust convex optimization[J]. Operations Research. 2014, 62(6):1358–1376.

[79] Isii K. On sharpness of tchebycheff-type inequalities[J]. Annals of the Institute of Statistical Mathematics. 1962, 14(1):185–197.

[80] Dhara A, Dutta J. Optimality conditions in convex optimization: a finite-dimensional view[M]. [S.l.]: CRC Press, 2011.

[81] Lagarias J C, Reeds J A, Wright M H, et al. Convergence properties of the nelder–mead simplex method in low dimensions[J]. Siam Journal on Optimization A Publication of the Society for Industrial & Applied Mathematics. 1997, 9(1):112–147.

[82] Zhai Q, Guan X, Cheng J, et al. Fast identification of inactive security constraints in scuc problems[J]. IEEE Transactions on Power Systems. 2010, 25(4):1946–1954.

[83] Gurobi Optimization I. Gurobi Optimizer Reference Manual[M/OL]. 2015. http://www.gurobi.com.

[84] Hu W, Min Y, Zhou Y, et al. Wind power forecasting errors modelling approach considering temporal and spatial dependence[J]. Journal of Modern Power Systems and Clean Energy. 2017, 5(3):489.

[85] Hodge B M, Lew D, Milligan M, et al. Wind Power Forecasting Error Distributions: An International Comparison[R]. [S.l.]: [s.n.], 2012.

[86] Lorca A, Sun X A. Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind[J]. IEEE Transactions on Power Systems. 2015, 30(4): 1702–1713.

[87] Bertsimas D, Gupta V, Kallus N. Data-driven robust optimization[J]. arXiv preprint arXiv:1401.0212. 2013.

[88] Duan C, Fang W, Jiang L, et al. Distributionally robust chance-constrained voltage-concerned dc-opf with wasserstein metric[J]. arXiv:1706.05538. 2017.

[89] Wang C, Liu F, Wang J, et al. Robust risk-constrained unit commitment with large-scale wind generation: An adjustable uncertainty set approach[J]. IEEE Transactions on Power Systems. 2017, 32 (1):723–733.

[90] Luo X, Wang J, Dooner M, et al. Overview of current development in electrical energy storage technologies and the application potential in power system operation[J]. Applied Energy. 2015, 137: 511–536.

[91] Wogrin S, Gayme D. Optimizing storage siting, sizing, and technology portfolios in transmission-constrained networks[J]. IEEE Transactions on Power Systems. Nov 2015, 30(6):3304–3313.

[92] Gill S, Kockar I, Ault G. Dynamic optimal power flow for active distribution networks[J]. IEEE Transactions on Power Systems. Jan 2014, 29(1):121–131.

[93] Gabash A, Li P. Active-reactive optimal power flow in distribution networks with embedded generation and battery storage[J]. IEEE Transactions on Power Systems. Nov 2012, 27(4):2026–2035.

[94] Gayme D, Topcu U. Optimal power flow with large-scale storage integration[J]. IEEE Transactions on Power Systems. May 2013, 28(2):709–717.

[95] Vrakopoulou M, Margellos K, Lygeros J, et al. A probabilistic framework for reserve scheduling and security assessment of systems with high wind power penetration[J]. IEEE Transactions on Power Systems. 2013, 28(4):3885–3896.

[96] Street A, Moreira A, Arroyo J M. Energy and reserve scheduling under a joint generation and transmission security criterion: An adjustable robust optimization approach[J]. IEEE Transactions on Power Systems. 2014, 29(1):3–14.

[97] Wei W, Liu F, Mei S, et al. Robust energy and reserve dispatch under variable renewable generation[J]. IEEE Transactions on Smart Grid. 2015, 6(1):369–380.

[98] Li Z, Wu W, Zhang B, et al. Adjustable robust real-time power dispatch with large-scale wind power integration[J]. IEEE transactions on Sustainable Energy. 2015, 6(2):357–368.

[99] Ding T, Bie Z, Bai L, et al. Adjustable robust optimal power flow with the price of robustness for large-scale power systems[J]. IET Generation, Transmission & Distribution. 2016, 10(1):164–174.

[100] Ye H, Li Z. Necessary conditions of line congestions in uncertainty accommodation[J]. IEEE Transactions on Power Systems. 2016, 31(5):4165–4166.

[101] Li Z, Guo Q, Sun H, et al. Sufficient conditions for exact relaxation of complementarity constraints for storage-concerned economic dispatch[J]. IEEE Transactions on Power Systems. 2016, 31(2): 1653–1654.

[102] Duan C, Jiang L, Fang W, et al. Improved sufficient conditions for exact convex relaxation of storage-concerned ed[J]. arXiv preprint arXiv:1603.07875. 2016.

[103] Yu H, Rosehart W. An optimal power flow algorithm to achieve robust operation considering load and renewable generation uncertainties[J]. IEEE Transactions on Power Systems. 2012, 27(4):1808–1817.

[104] PSERC. Challenges in integrating renewable technologies into an electric power system[J]. White Paper. 2010.

[105] Jabr R. Adjustable robust opf with renewable energy sources[J]. IEEE Transactions on Power Systems. Nov 2013, 28(4):4742–4751.

[106] Duan C, Jiang L, Fang W, et al. Data-driven affinely adjustable distributionally robust unit commitment[J]. IEEE Transactions on Power Systems. 2017, PP(99):1–1.

[107] Roald L, Oldewurtel F, Van Parys B, et al. Security constrained optimal power flow with distributionally robust chance constraints[J]. arXiv preprint arXiv:1508.06061. 2015.

[108] Xie W, Ahmed S. Distributionally robust chance constrained optimal power flow with renewables: A conic reformulation[J]. optimization-online. 2016.

[109] Mohajerin Esfahani P, Kuhn D. Data-driven distributionally robust optimization using the wasserstein metric: Performance guarantees and tractable reformulations[J]. arXiv preprint arXiv:1512.05489. 2015.

[110] Gao R, Kleywegt A J. Distributionally robust stochastic optimization with wasserstein distance[J]. arXiv preprint arXiv:1604.02199. 2016.

[111] Kundur P, Balu N J, Lauby M G. Power system stability and control: volume 7[M]. [S.l.]: McGraw-hill New York, 1994.

[112] Yang J, Zhang N, Kang C, et al. A state-independent linear power flow model with accurate estimation of voltage magnitude[J]. IEEE Transactions on Power Systems. 2016, PP(99):1–1.

[113] Zhao C, Guan Y. Data-driven risk-averse stochastic optimization with wasserstein metric[J]. Available on optimization online. 2015.

[114] Zhao C, Guan Y. Data-driven risk-averse two-stage stochastic program with ζ-structure probability metrics[J]. Available on Optimization Online. 2015.

[115] Bolley F, Villani C. Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities[J]. Ann. Fac. Sci. Toulouse Math. (6). 2005, 14(3):331–352.

[116] Zimmerman R D, Murillo-SáNchez C E, Thomas R J. Matpower: Steady-state operations, planning, and analysis tools for power systems research and education[J]. IEEE Transactions on Power Systems. 2011, 26(1):12–19.

[117] Zimmerman R D, Murillo-Sánchez C E, Thomas R J. Matpower's extensible optimal power flow architecture[C]// IEEE. Power & Energy Society General Meeting, 2009. PES'09. IEEE. [S.l.]: IEEE, 2009: 1–7.

[118] Lofberg J. Yalmip: A toolbox for modeling and optimization in matlab[C]// IEEE. Computer Aided Control Systems Design, 2004 IEEE International Symposium on. [S.l.]: IEEE, 2004: 284–289.

[119] Byrd R H, Nocedal J, Waltz R A. KNITRO: An integrated package for nonlinear optimization[M]. [S.l.]: Springer, 2006: 35–59.

[120] Lee D, Baldick R. Probabilistic wind power forecasting based on the laplace distribution and golden search[C]. Ieee/pes Transmission and Distribution Conference and Exposition. [S.l.], 2016: 1-5.

[121] Aoki K, Fan M, Nishikori A. Optimal var planning by approximation method for recursive mixed-integer linear programming[J]. IEEE Transactions on Power Systems. 1988, 3(4):1741–1747.

[122] Yang G Y, Hovland G, Majumder R, et al. Tcsc allocation based on line flow based equations via mixed-integer programming[J]. IEEE Transactions on Power Systems. 2007, 22(4):2262–2269.

[123] Hsiao Y T, Liu C C, Chiang H D, et al. A new approach for optimal var sources planning in large scale electric power systems[J]. IEEE Transactions on Power Systems. 1993, 8(3):988–996.

[124] Chattopadhyay D, Bhattacharya K, Parikh J. Optimal reactive power planning and its spot-pricing: an integrated approach[J]. IEEE Transactions on Power Systems. 1995, 10(4):2014–2020.

[125] Dong F, Chowdhury B H, Crow M L, et al. Improving voltage stability by reactive power reserve management[J]. IEEE Transactions on Power Systems. 2005, 20(1):338–345.

[126] Mínguez R, Milano F, Zárate-Miñano R, et al. Optimal network placement of svc devices[J]. IEEE Transactions on Power Systems. 2007, 22(4):1851–1860.

[127] Leonidaki E, Georgiadis D, Hatziargyriou N. Decision trees for determination of optimal location and rate of series compensation to increase power system loading margin[J]. IEEE Transactions on Power Systems. 2006, 21(3):1303–1310.

[128] Paterni P, Vitet S, Bena M, et al. Optimal location of phase shifters in the french network by genetic algorithm[J]. IEEE Transactions on Power Systems. 1999, 14(1):37–42.

[129] Ghahremani E, Kamwa I. Optimal placement of multiple-type facts devices to maximize power system loadability using a generic graphical user interface[J]. IEEE Transactions on Power Systems. 2013, 28(2):764–778.

[130] Gomez T, Perez-Arriaga I, Lumbreras J, et al. A security-constrained decomposition approach to optimal reactive power planning[J]. IEEE Transactions on Power Systems. 1991, 6(3):1069–1076.

[131] Granville S, Lima M A. Application of decomposition techniques to var planning: methodological and computational aspects[J]. IEEE Transactions on Power Systems. 1994, 9(4):1780–1787.

[132] Lima F G, Galiana F D, Kockar I, et al. Phase shifter placement in large-scale systems via mixed integer linear programming[J]. IEEE Transactions on Power Systems. 2003, 18(3):1029–1034.

[133] Bach F, Jenatton R, Mairal J, et al. Optimization with sparsity-inducing penalties[J]. Foundations and Trends® in Machine Learning. 2012, 4(1):1–106.

[134] Jabr R A, Martins N, Pal B C, et al. Contingency constrained var planning using penalty successive conic programming[J]. IEEE Transactions on Power Systems. 2012, 27(1):545–553.

[135] Beck A, Eldar Y C. Sparsity constrained nonlinear optimization: Optimality conditions and algorithms[J]. SIAM Journal on Optimization. 2013, 23(3):1480–1509.

[136] Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundations and Trends® in Machine Learning. 2011, 3(1):1–122.

[137] Torres G L, Quintana V H. On a nonlinear multiple-centrality-corrections interior-point method for optimal power flow[J]. IEEE Transactions on Power Systems. 2001, 16(2):222–228.

[138] Wu Y C, Debs A S, Marsten R E. A direct nonlinear predictor-corrector primal-dual interior point algorithm for optimal power flows[J]. IEEE Transactions on Power Systems. 1994, 9(2):876–883.

[139] Yang A Y, Sastry S S, Ganesh A, et al. Fast ℓ 1-minimization algorithms and an application in robust face recognition: A review[C]// IEEE. Image Processing (ICIP), 2010 17th IEEE International Conference on. [S.l.]: IEEE, 2010: 1849–1852.

[140] Xu Z, Chang X, Xu F, et al. $l\_\{1/2\}$ regularization: A thresholding representation theory and a fast solver[J]. IEEE Transactions on neural networks and learning systems. 2012, 23(7):1013–1027.

[141] Cao W, Sun J, Xu Z. Fast image deconvolution using closed-form thresholding formulas of regularization[J]. Journal of visual communication and image representation. 2013, 24(1):31–41.

[142] Blumensath T, Davies M E. Iterative thresholding for sparse approximations[J]. Journal of Fourier Analysis and Applications. 2008, 14(5):629–654.

[143] Natarajan B K. Sparse approximate solutions to linear systems[J]. SIAM journal on computing. 1995, 24(2):227–234.

[144] Zong-Ben, Hai-Liang, WANG, et al. Representative of l1/2 regularization among lq (0 〈q ≤ 1) regularizations: an experimental study based on phase diagram[J]. Acta Automatica Sinica. 2012, 38 (7):1225–1228.

[145] Yang J, Zhang Y. Alternating direction algorithms for \ell_1-problems in compressive sensing[J]. SIAM journal on scientific computing. 2011, 33(1):250–278.

[146] Xu Y, Yin W, Wen Z, et al. An alternating direction algorithm for matrix completion with nonnegative factors[J]. Frontiers of Mathematics in China. 2012, 7(2):365–384.

[147] Jiang B, Ma S, Zhang S. Alternating direction method of multipliers for real and complex polynomial optimization models[J]. Optimization. 2014, 63(6):883–898.

[148] Erseghe T. Distributed optimal power flow using admm[J]. IEEE Transactions on Power Systems. 2014, 29(5):2370–2380.

[149] Phan D, Kalagnanam J. Some efficient optimization methods for solving the security-constrained optimal power flow problem[J]. IEEE Transactions on Power Systems. 2014, 29(2):863–872.

[150] Zhang X P, Rehtanz C, Pal B. Flexible AC transmission systems: modelling and control[M]. [S.l.]: Springer Science & Business Media, 2012.

[151] Zhang X P. Advanced modeling of the multicontrol functional static synchronous series compensator (sssc) in newton power flow[J]. IEEE Transactions on Power Systems. 2003, 18(4):1410–1416.

[152] An S, Condren J, Gedra T W. An ideal transformer upfc model, opf first-order sensitivities, and application to screening for optimal upfc locations[J]. IEEE Transactions on Power Systems. 2007, 22(1):68–75.

[153] Bertsekas D P. Nonlinear programming[M]. [S.l.]: Athena scientific Belmont, 1999.

[154] Lavaei J, Low S H. Zero duality gap in optimal power flow problem[J]. IEEE Transactions on Power Systems. 2012, 27(1):92–107.

[155] Bukhsh W A, Grothey A, McKinnon K I, et al. Local solutions of the optimal power flow problem[J]. IEEE Transactions on Power Systems. 2013, 28(4):4780–4788.

[156] Bach F, Jenatton R, Mairal J, et al. Structured sparsity through convex optimization[J]. Statistical Science. 2012, 27(4):450–468.

[157] Taylor J A. Convex optimization of power systems[M]. [S.l.]: Cambridge University Press, 2015.

[158] Gan L, Li N, Topcu U, et al. Exact convex relaxation of optimal power flow in radial networks[J]. IEEE Transactions on Automatic Control. 2015, 60(1):72–87.

[159] Madani R, Sojoudi S, Lavaei J. Convex relaxation for optimal power flow problem: Mesh networks[J]. IEEE Transactions on Power Systems. 2015, 30(1):199–211.

[160] Jabr R A. Exploiting sparsity in sdp relaxations of the opf problem[J]. IEEE Transactions on Power Systems. 2012, 27(2):1138–1139.

[161] Molzahn D K, Holzer J T, Lesieutre B C, et al. Implementation of a large-scale optimal power flow solver based on semidefinite programming[J]. IEEE Transactions on Power Systems. 2013, 28(4): 3987–3998.

[162] Lasserre J B. Global optimization with polynomials and the problem of moments[J]. SIAM Journal on Optimization. 2001, 11(3):796–817.

[163] Nie J. Optimality conditions and finite convergence of lasserre’s hierarchy[J]. Mathematical programming. 2014, 146(1-2):97–121.

[164] Ghaddar B, Marecek J, Mevissen M. Optimal power flow as a polynomial optimization problem[J]. IEEE Transactions on Power Systems. 2015, PP(99):1–8.

[165] Josz C, Maeght J, Panciatici P, et al. Application of the moment-sos approach to global optimization of the opf problem[J]. IEEE Transactions on Power Systems. Jan 2015, 30(1):463–470.

[166] Molzahn D, Hiskens I. Sparsity-exploiting moment-based relaxations of the optimal power flow problem[J]. IEEE Transactions on Power Systems. Nov 2015, 30(6):3168–3180.

[167] Tamura Y, Mori H, Iwamoto S. Relationship between voltage instability and multiple load flow solutions in electric power systems[J]. IEEE Transactions on power apparatus and systems. 1983, PAS-102(5):1115–1125.

[168] Wang X F, Song Y, Irving M. Modern power systems analysis[M]. [S.l.]: Springer Science & Business Media, 2010.

[169] Blekherman G, Parrilo P A, Thomas R R. Semidefinite optimization and convex algebraic geometry[M]. [S.l.]: SIAM, 2012.

[170] Lasserre J B. Moments, positive polynomials and their applications[M]. [S.l.]: World Scientific, 2009.

[171] Kim S, Kojima M, Waki H. Generalized lagrangian duals and sums of squares relaxations of sparse polynomial optimization problems[J]. SIAM Journal on Optimization. 2005, 15(3):697–719.

[172] Waki H, Kim S, Kojima M, et al. Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity[J]. SIAM Journal on Optimization. 2006, 17(1):218–242.

[173] Lasserre J B. Convergent sdp-relaxations in polynomial optimization with sparsity[J]. SIAM Journal on Optimization. 2006, 17(3):822–843.

[174] Blair J R, Peyton B. An introduction to chordal graphs and clique trees[M]. [S.l.]: Springer, 1993: 1–29.

[175] Löfberg J. Yalmip : A toolbox for modeling and optimization in MATLAB[C/OL]. Proceedings of the CACSD Conference. Taipei, Taiwan, 2004. `http://users.isy.liu.se/johanl/yalmip`.

[176] Mosek A. The mosek optimization toolbox for matlab manual[J]. Version 7.1 (Revision 28). 2015, 17.

[177] Gopalakrishnan A, Raghunathan A, Nikovski D, et al. Global optimization of multi-period optimal power flow[C]. American Control Conference (ACC), 2013. [S.l.], June 2013: 1157-1164.

[178] Castillo A, Gayme D. Profit maximizing storage allocation in power grids[C]. Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on. [S.l.], Dec 2013: 429-435.

[179] Jabr R. Exploiting sparsity in sdp relaxations of the opf problem[J]. IEEE Transactions on Power Systems. May 2012, 27(2):1138–1139.

[180] Molzahn D, Holzer J, Lesieutre B, et al. Implementation of a large-scale optimal power flow solver based on semidefinite programming[J]. IEEE Transactions on Power Systems. Nov 2013, 28(4): 3987–3998.

[181] Skajaa A, Andersen E D, Ye Y. Warmstarting the homogeneous and self-dual interior point method for linear and conic quadratic problems[J]. Mathematical Programming Computation. 2013, 5(1): 1–25.

[182] Goldstein T, O'Donoghue B, Setzer S, et al. Fast alternating direction optimization methods[J]. SIAM Journal on Imaging Sciences. 2014, 7(3):1588–1623.

[183] Sönmez S, Ayasun S, Nwankpa C O. An exact method for computing delay margin for stability of load frequency control systems with constant communication delays[J]. IEEE Transactions on Power Systems. 2016, 31(1):370–377.

[184] Michiels W, Niculescu S I. Stability, Control, and Computation for Time-Delay Systems: An Eigenvalue-Based Approach: volume 27[M]. [S.l.]: SIAM, 2014.

[185] Milano F. Small-signal stability analysis of large power systems with inclusion of multiple delays[J]. IEEE Transactions on Power Systems. 2016, 31(4):3257–3266.

[186] Jia H, Yu X. A simple method for power system stability analysis with multiple time delays[C]// IEEE. Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century. [S.l.]: IEEE, 2008: 1–7.

[187] Ayasun S, Eminoğlu U, Sönmez Ş. Computation of stability delay margin of time-delayed generator excitation control system with a stabilizing transformer[J]. Mathematical Problems in Engineering. 2014, 2014.

[188] Sönmez S, Ayasun S, Eminoglu U. Computation of time delay margins for stability of a single-area load frequency control system with communication delays[J]. WSEAS Trans. Power Syst. 2014, 9: 67–76.

[189] Sönmez S, Ayasun S. Stability region in the parameter space of pi controller for a single-area load frequency control system with time delay[J]. IEEE Transactions on Power Systems. 2016, 31(1): 829–830.

[190] Thangaiah J M, Parthasarathy R. Delay-dependent stability analysis of power system considering communication delays[J]. International Transactions on Electrical Energy Systems. 2016.

[191] Ramakrishnan K, Ray G. Stability criteria for nonlinearly perturbed load frequency systems with time-delay[J]. IEEE Journal on Emerging and Selected Topics in Circuits and Systems. 2015, 5(3): 383–392.

[192] Ahmadi A, Aldeen M. An lmi approach to the design of robust delay-dependent overlapping load frequency control of uncertain power systems[J]. International Journal of Electrical Power & Energy Systems. 2016, 81:48–63.

[193] Peng C, Zhang J. Delay-distribution-dependent load frequency control of power systems with probabilistic interval delays[J]. IEEE Transactions on Power Systems. 2016, 31(4):3309–3317.

[194] Mason R P, Papachristodoulou A. Chordal sparsity, decomposing sdps and the lyapunov equation[C]// IEEE. American Control Conference (ACC), 2014. [S.l.]: IEEE, 2014: 531–537.

[195] Park P, Ko J W, Jeong C. Reciprocally convex approach to stability of systems with time-varying delays[J]. Automatica. 2011, 47(1):235–238.

[196] Kim S, Kojima M, Mevissen M, et al. Exploiting sparsity in linear and nonlinear matrix inequalities via positive semidefinite matrix completion[J]. Mathematical programming. 2011, 129(1):33–68.

[197] Artin M. Algebra[M/OL]: Prentice Hall, 1991. https://books.google.co.uk/books?id=C_juAAAAMAAJ.

[198] Tütüncü R H, Toh K C, Todd M J. Solving semidefinite-quadratic-linear programs using sdpt3[J]. Mathematical programming. 2003, 95(2):189–217.

[199] Athay T, Podmore R, Virmani S. A practical method for the direct analysis of transient stability[J]. IEEE Transactions on Power Apparatus and Systems. 1979, PAS-98(2):573–584.

[200] Nakata K, Fujisawa K, Fukuda M, et al. Exploiting sparsity in semidefinite programming via matrix completion ii: Implementation and numerical results[J]. Mathematical Programming. 2003, 95(2): 303–327.

[201] Rerkpreedapong D, Hasanovic A, Feliachi A. Robust load frequency control using genetic algorithms and linear matrix inequalities[J]. IEEE Transactions on Power Systems. 2003, 18(2):855–861.

[202] Pal B, Chaudhuri B. Robust control in power systems[M]. [S.l.]: Springer Science & Business Media, 2006.

[203] Li Y, Yang D, Liu F, et al. Interconnected Power Systems: Wide-area Dynamic Monitoring and Control Applications[M]. [S.l.]: Springer, 2015.

[204] Gong C, Zhu G, Wu L. New weighted integral inequalities and its application to exponential stability analysis of time-delay systems[J]. IEEE Access. 2016, 4:6231–6237.

[205] Haddad W M, Chellaboina V. Nonlinear dynamical systems and control: a Lyapunov-based approach[M]. [S.l.]: Princeton University Press, 2008.

[206] Li C, Du Z, Ni Y, et al. Reduced model-based coordinated design of decentralized power system controllers[J]. IEEE Transactions on Power Systems. 2016, 31(3):2172–2181.

[207] Antoulas A C. Approximation of Large-Scale Dynamical Systems[M]. [S.l.]: Society for Industrial and Applied Mathematics, 2005: xxvi.

[208] Kang-Zhi Liu Y Y. Robust Control: Theory and Applications[M]. [S.l.]: John Wiley & Sons, Inc., 2016.

[209] Hassibi A, How J, Boyd S. A path-following method for solving bmi problems in control[C]. Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251): volume 2. [S.l.], 1999: 1385-1389 vol.2.

[210] Ostertag E. An improved path-following method for mixed controller design[J]. IEEE Transactions on Automatic Control. 2008, 53(8):1967–1971.

[211] Canizares C, Fernandes T, Geraldi E, et al. Benchmark models for the analysis and control of small-signal oscillatory dynamics in power systems[J]. IEEE Transactions on Power Systems. 2016, 32(1): 715–722.

[212] Canizares C, Fernandes T, Geraldi Jr E, et al. Benchmark systems for small signal stability analysis and control[J]. http://resourcecenter. ieee-pes. org/pes/product/technical-reports/PESTR18. 2015, (PES-TR).

[213] Michiels W, Niculescu S I. Stability and Stabilization of Time-Delay Systems (Advances in Design & Control) (Advances in Design and Control)[M]. [S.l.]: Society for Industrial and Applied Mathematics, 2007.

[214] Mosek A. The mosek optimization toolbox for matlab manual[J]. Version 7.1 (Revision 28). 2015, 17.

[215] Martins N, Pinto H J C P, Lima L T G. Efficient methods for finding transfer function zeros of power systems[J]. IEEE Transactions on Power Systems. 1992, 7:3(3):1350–1361.

[216] Qiu Z, Deconinck G, Belmans R. A literature survey of optimal power flow problems in the electricity market context[C]. Power Systems Conference and Exposition, 2009. PSCE '09. IEEE/PES. [S.l.], 2009: 1-6.

[217] Granville S. Optimal reactive dispatch through interior point methods[J]. Power Systems IEEE Transactions on. 1994, 9(1):136–146.

[218] Capitanescu F, Wehenkel L. Experiments with the interior-point method for solving large scale optimal power flow problems[J]. Electric Power Systems Research. 2013, 95(2):276–283.

[219] Duan C, Fang W, Jiang L, et al. Facts devices allocation via sparse optimization[C]. Power and Energy Society General Meeting. [S.l.], 2016: 1-1.

[220] Nocedal J, Chter A, Waltz R A. Adaptive barrier update strategies for nonlinear interior methods[J]. Siam Journal on Optimization. 2010, 19(4):1674–1693.

[221] Wächter A, Biegler L T. Line search filter methods for nonlinear programming: Motivation and global convergence.[J]. Siam Journal on Optimization. 2006, 16(1):1–31.

[222] Wachter A, Biegler L T. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming[J]. Mathematical programming. 2006, 106(1):25–57.

[223] Fletcher R, Leyffer S. Nonlinear programming without a penalty function[J]. Mathematical Programming. 2002, 91(2):239–269.

[224] Bertsekas D P. Projected newton methods for optimization problems with simple constraints[C]. Decision and Control Including the Symposium on Adaptive Processes, 1981 IEEE Conference on. [S.l.], 2007: 762-767.

[225] Girard T D, Ware L B, Bernard G R, et al. Interior-point based algorithms for the solution of optimal power flow problems[J]. Electric Power Systems Research. 2007, 77(5−6):508–517.

[226] Mehrotra S. On the implementation of a primal-dual interior point method[J]. American Journal of Digestive Diseases. 1992, 2(4):575–601.

[227] Gondzio J. Multiple centrality corrections in a primal-dual method for linear programming[J]. Computational Optimization & Applications. 1996, 6(2):137–156.

Publications

Published

[1] **Duan C**, Fang W, Jiang L, et al. Distributionally Robust Chance-Constrained Approximate AC-OPF with Wasserstein Metric[J]. IEEE Transactions on Power Systems. 2018, PP(99):1–1.

[2] **Duan C**, Jiang L, Fang W, J. Liu. Data-driven affinely adjustable distributionally robust unit commitment[J]. IEEE Transactions on Power Systems. 2017, PP(99):1–1.

[3] **Duan C**, Zhang C K, Jiang L, Fang W, W. Yao. Structure-exploiting delay-dependent stability analysis applied to power system load frequency control[J]. IEEE Transactions on Power Systems. Nov 2017, 32(6): 4528–4540. SCI: FK1NE.

[4] **Duan C**, Jiang L, Fang W, J. Liu. Moment-sos approach to interval power flow[J]. IEEE Transactions on Power Systems. Jan 2017, 32(1):522–530. SCI: EH4KW.

[5] **Duan C**, Jiang L, Fang W, J. Liu. Data-driven distributionally robust energy-reserve-storage dispatch[J]. IEEE Transactions on Industrial Informatics. 2017, PP(99):1–1.

[6] **Duan C**, Fang W, Jiang L, S. Niu. FACTS devices allocation via sparse optimization[J]. IEEE Transactions on Power Systems. March 2016, 31(2):1308–1319. SCI: DG4CH.

[7] **Duan C**, Fang W, Jiang L, J. Liu. Adaptive barrier filter-line-search interior point method for optimal power flow with facts devices[J]. IET Generation, Transmission & Distribution. 2015, 9(16):2792–2798. SCI: CW9RC.

[8] **Duan C**, Jiang L, Fang W, J. Liu. Multi-period opf with energy storages and renewable sources: A parallel moment approach[C]. 2016 IEEE Power and Energy Society General Meeting (PESGM). [S.l.], July 2016: 1-5.

[9] Fan H, **Duan C**, Zhang C K, Jiang L, C. Mao, D. Wang. ADMM-based Multiperiod Optimal Power Flow Considering Plug-in Electric Vehicles Charging [J]. IEEE Transactions on Power Systems. 2017, PP.

[10] Du Y, Jiang L, **Duan C**, Y. Li, J. S. Smith. Energy consumption scheduling of hvac considering weather forecast error through distributionally robust approach[J]. IEEE Transactions on Industrial Informatics. 2017, PP(99): 1–1.

[11] Yao L, Wang X, **Duan C**, et al. Data-driven distributionally robust reserve and energy scheduling over wasserstein balls[J]. IET Generation, Transmission & Distribution. 2017, PP.

Pending

[1] **Duan C**, Li C, Jiang, L, Fang W, et al. Path-following Method to Design Power System Damping Controllers Considering Time Delays and Model Reduction Errors[J]. working paper.